

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run On: May 12, 2003, 10:09:12 ; Search time 78 Seconds  
(without alignments)  
736.295 Million cell updates/sec

Title: US-10-057-531A-2

Perfect score: 2275

Sequence: 1 MHHHHHHSSGLVPRGSMKE.....TCECTKPDVPLFDGIFCSS 431

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002.\*

1:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23:	/SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1970	86.6	375	22 AAB83926	A major merozoite
2	1175	51.6	1639	19 AAW54145	P. falciparum synt.
3	1161.5	51.1	355	20 AAY09372	Merozoite surface
4	1161.5	51.1	355	20 AAY05832	Merozoite surface
5	1161.5	51.1	361	20 AAY09373	Merozoite surface
6	1161.5	51.1	361	20 AAY05833	Merozoite surface
7	1154.5	50.7	376	20 AAY09374	Modified merozoite
8	1154.5	50.7	376	20 AAY05834	Modified merozoite
9	1007.5	44.3	1654	6 AAP50777	Sequence of the PI
10	640.5	28.2	262	22 AAB37610	Merozoite surface

11	527	23.2	116	18 AAW36103	PFMSPI(p19)A prote
12	527	23.2	116	18 AAW22592	PFMSPI(p19)A prote
13	527	23.2	127	18 AAW22593	PFMSPI(p19)S prote
14	527	23.2	127	18 AAW36102	PFMSPI(p19)S prote
15	525	23.1	108	22 AAB37609	Merozoite surface
16	524	23.0	96	22 AAB37608	Merozoite surface
17	406	17.8	206	22 AAU00669	Plasmodium vivax m
18	302	13.3	402	23 ABG67266	Polythoa 2 fluore
19	301	13.2	407	23 ABG67267	Polythoa 2 fluore
20	297	13.1	407	23 ABG67267	Polythoa 2 fluore
21	294	12.9	54	14 AAR41357	MSPIEGF2B EGF2-11k
22	283.5	12.5	1807	22 AAB85697	Recombinant protei
23	283.5	12.5	2028	22 AAB85698	Recombinant protei
24	282	12.4	49	14 AAR41355	MSPIEGF1B EGF1-11k
25	280	12.3	108	22 AAB47487	PV200C polypeptide
26	279	12.3	49	14 AAR41354	MSPIEGF1A EGF1-11k
27	278.5	12.2	289	22 AAE13033	Helicobacter pylor
28	278.5	12.2	289	23 AAB47732	HOP38 #5. Helicob
29	278.5	12.2	824	23 AAM48247	Thioredoxin/deoxyx
30	278	12.2	54	14 AAR41356	MSPIEGF2A EGF2-11k
31	277.5	12.2	692	21 AAB11532	SEN virus protein
32	275	12.1	467	21 AAY85150	Mouse secreted pro
33	275	12.1	467	21 AAY79067	Murine secreted pr
34	274	12.0	76	21 AAB26110	H. contortus clone
35	274	12.0	167	22 AAB85253	Thioredoxin functi
36	274	12.0	167	22 AAB85254	Thioredoxin functi
37	274	12.0	551	22 AAB85251	Plant thioredoxin-
38	274	12.0	551	22 AAB85252	Plant thioredoxin-
39	272	12.0	93	22 AAB37611	Merozoite surface
40	263.5	11.6	138	21 AAB11534	SEN virus protein
41	259.5	11.4	408	20 AAW84315	TrxA-rabbit tissue
42	256.5	11.3	106	14 AAR41358	P. yoelii combined
43	255.5	11.2	284	23 AAE20460	Recombinant (SPKR)
44	250	11.0	59	23 ABB80661	N-His botulinum to
45	250	11.0	218	21 AAB11533	SEN virus protein

#### ALIGNMENTS

RESULT 1  
AAB83926  
ID AAB83926 standard; Protein; 375 AA.  
XX  
AC AAB83926;  
XX  
DT 23-JUL-2001 (first entry)  
XX  
DE A major merozoite surface protein-1 fragment of 42kDa.  
XX  
KW Major merozoite surface protein-1; MSPI-42; melittin signal peptide;  
KW malaria vaccine  
XX  
US Plasmodium falciparum.  
XX  
PN WO200134188-A1.  
XX  
PD 17-MAY-2001.  
XX  
PP 09-NOV-2000; 2000WO-US31064.  
XX  
PR 12-NOV-1999; 99US-0165178.  
PR 01-DEC-1999; 99US-0168327.  
PR 22-AUG-2000; 2000US-0226861.  
XX  
PA (UYHA-) UNIV HAWAII.  
PA (UYCH-) UNIV CHINESE HONG KONG.  
PA (QUEE-) QUEEN EMMA FOUND.  
XX  
PI Hui GSN, Lap-Yin P, Ho WKK;  
XX  
DR WPI; 2001-335879/35.  
DR N-PSDB; AAF89840.

XX Producing malaria vaccine, useful for treatment or prevention of all  
PT forms of malaria in humans, by expressing immunogenic merozoite protein  
PT fragment in a baculovirus system  
XX  
XX  
PS Example 3; Page 87-88; 95pp; English.  
XX  
XX The present sequence represents a major merozoite surface protein-1  
CC C-terminal fragment of 42kDa (MSP-1-42). This fragment is linked to a  
CC melittin signal peptide, and then expressed in a in a  
CC silkworm/baculovirus system. The protein is used to prepare a  
CC malaria vaccine, which is used to treat or prevent malaria, caused by  
CC any of the four species of Plasmodium that infect humans.  
XX  
XX Sequence 375 AA;  
SQ  
Query Match 86.6%; Score 1970; DB 22; Length 375;  
Best Local Similarity 100.0%; Pred. No. 2e-132;  
Matches 375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 57 MAISVTMDNLSGPFENEVDVYILKPLAGVYRSLSKQIEKNFTFTNLDNLSRLKRRK 116  
Db  
QY 177 YFLOVLESDLMQFKHISSENYIEDSFLLNSEQNTLLSKYIKESVENDIKFAQEG 176  
Db 61 YFLOVLESDLMQFKHISSENYIEDSFLLNSEQNTLLSKYIKESVENDIKFAQEG 120  
QY 177 SYEYKVLAKYKDDLESIRKVIKESKEKPPSPPTPPSPAKTDEQKESKFLPFTLN 236  
Db 121 SYEYKVLAKYKDDLESIRKVIKESKEKPPSPPTPPSPAKTDEQKESKFLPFTLN 180  
QY 237 LYNNLVNKIDYLLNLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNPYDFA 296  
Db 181 LYNNLVNKIDYLLNLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNPYDFA 240  
QY 297 KLINDTKKMLGKLLSTGLVQNFPTTIISKLIEGKFQDMLNISOHOCVKKQCPENSGCF 356  
Db 241 KLINDTKKMLGKLLSTGLVQNFPTTIISKLIEGKFQDMLNISOHOCVKKQCPENSGCF 300  
QY 357 RHLDERECKCLLNYKQGDCKVCNPNPTCNENNGCGDADATCTEEDSGSRKKITCECT 416  
Db 301 RHLDERECKCLLNYKQGDCKVCNPNPTCNENNGCGDADATCTEEDSGSRKKITCECT 360  
QY 417 KPDSTPLFDGIFCSS 431  
Db 361 KPDSTPLFDGIFCSS 375

RESULT 2  
AAW54145  
ID AAW54145 standard; Protein; 1639 AA.  
XX  
AC AAW54145;  
XX  
XX 23-SEP-1998 (first entry)  
XX  
DE P. falciparum synthetic gpl90 protein.  
XX  
XX gpl90; malaria; MSP-1; merozoite surface protein; stability; vaccine;  
KW monoclonal antibody; passive immunisation; parasite.  
XX  
XX Plasmodium falciparum.  
OS Synthetic.  
XX  
XX WO9814583-A2.  
XX  
XX 09-APR-1998.  
XX  
XX 02-OCT-1997; 97WO-EP05441.  
XX  
XX 02-OCT-1996; 96DE-4040817.  
XX

PA (BUJA/) BUJARD H.  
XX Bujard H, Pan W, Tolle R;  
XX WPI: 1998-240088/21.  
DR N-PSDB; AAV21451, AAV35363.  
XX  
XX Recombinant production of complete gpl90/MSP-1 Plasmodium surface  
PT protein - useful in anti-malaria vaccines, also stabilising genes by  
PT reducing their AT content  
XX  
XX Example 1; Fig 3c; 48pp; German.  
XX  
XX This sequence represents a modified Plasmodium falciparum gpl90/MSP-1  
CC (merozoite surface) protein. The gene encoding this protein has been  
CC stabilised by reducing the AT content of the nucleotide sequence. Such a  
CC protein is useful in vaccines against malaria or for producing monoclonal  
CC antibodies (for passive immunisation). The complete gpl90 protein can now  
CC be produced outside the parasite and has, at least over extended regions,  
CC the native pattern of folding. Larger amounts of the protein can be  
CC produced recombinantly than would be possible using the parasites as  
CC source.  
XX  
XX Sequence 1639 AA;  
SQ  
Query Match 51.6%; Score 1175; DB 19; Length 1639;  
Best Local Similarity 56.6%; Pred. No. 7.3e-75;  
Matches 226; Conservative 59; Mismatches 90; Indels 24; Gaps 4;  
QY 34 PDLGTTDDDDKAMADIGSIEGRGTMAISVTMDNLSGPFENEVDVYILKPLAGVYRSLSKQ 93  
Db 1244 PIFGESEED--YDGLQGVVTGEAVTPSV-IDNLSKIENEYEVLYLPLAGVYRSLSKQ 1300  
QY 94 EKNFTFTNLDNLSRLKRRKYFLDVLESDLMQFKHISSENYIEDSFLLNSEQNT 153  
Db 1301 ENNVYTFNVNVDILNSRFNKNRENFKNVLESDDLIPYKDLTSSNVVVDYKFLNKRDK 1360  
QY 154 LLKYIKYIKESVENDIKFAQEGISYIEKVLAKYKDDLESIRKVIKESKEKPPSPPTTP 213  
Db 1361 FLSSYNTIKDSIDTDINFANDVLGYIKLSEKYSKSLDSIKKYI----- 1404  
QY 214 SPAKTDEQKESKFLPFTLNFTNLTNLYNNLVNKIDYLLNLKAKINDCNVEKDEAHVKIT 273  
Db 1405 ----NDKQGENEKYLPFLNFTNLTNLYNNLVNKIDYLLNLKAKINDCNVEKDEAHVKIT 1460  
QY 274 SLDKAIDDKIDLFKNPYDFAIKKLLINDDTKDKMLGKLLSTGLV-QNFPNTIISKLEGG 332  
Db 1461 NYLKTIQDKLADFKNNNFVGIADLSTDYNNHNNLLTFLSTGMVFENLAKTVLSNLLDGN 1520  
QY 333 FQDMLNISOHOCVKKQCPENSGCFRHLDERECKCLLNYKQGDCKVCNPNPTCNENNGG 392  
Db 1521 LQGLMLNISOHOCVKKQCPENSGCFRHLDERECKCLLNYKQGDCKVCNPNPTCNENNGG 1580  
QY 393 CDADATCTEEDSGSRKKITCECTKPDSTPLFDGIFCSS 431  
Db 1581 CDADAKTEEDSGSGNKKITCECTKPDSTPLFDGIFCSS 1619

RESULT 3  
AAW09372  
ID AAW09372 standard; Protein; 355 AA.  
XX  
AC AAW09372;  
XX  
XX 31-AUG-1999 (first entry)  
XX  
XX Merozoite surface protein MSP-1-42.  
XX  
XX MSP-1; merozoite surface protein; malaria; vaccine;  
KW protein engineering; protein expression; codon usage;  
XX transgenic animal.  
XX  
XX Plasmodium falciparum.  
OS



```
XX WO9920774-A2.
PN
PD
PD 29-APR-1999.
XX
PF 20-OCT-1998; 98WO-US222226.
XX
PR 15-MAY-1998; 98US-0085649.
PR 20-OCT-1997; 97US-0062592.
XX
PA (GENZ ) GENZYME TRANSGENICS CORP.
XX
PI Chen LH, Meade H;
XX
DR WPI: 1999-288313/24.
DR P-PSDB; AAX56008.
XX
PT Modified malarial protein for use in anti-malarial vaccines
XX
PS Example; Fig 1; 35pp; English.
XX
CC The present sequence represents a 42 kDa C-terminal portion of
CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC important target for the development of a vaccine against
CC Plasmodium falciparum. The nucleic acid sequence encoding MSP-1-42
CC has been modified (see AAX56008) compared to the native sequence (see
CC AAX56009) such that 306 nucleotide positions have been replaced to
CC lower the AT content (from 76 to 49.7%) and to eliminate 10 mRNA
CC instability motifs while maintaining the same protein amino acid
CC sequence. These alterations allow MSP-1-42 to be expressed in
CC mammalian cell culture and in transgenic mice. Native MSP-1-12
CC is known to be difficult to express in cell culture systems,
CC mammalian cell culture systems or in transgenic animals. The
CC invention allows expression of MSP-1 protein in the milk of
CC transgenic animals, and also provides a DNA vaccine comprising a
CC vector containing the altered MSP-1-42 sequence.
XX
SQ Sequence 355 AA;
Query Match 51.1%; Score 1161.5; DB 20; Length 355;
Best Local Similarity 59.1%; Pred. No. 8.8e-75;
Matches 218; Conservative 54; Mismatches 76; Indels 21; Gaps 2;
QY 63 MNILSGFENEVDVYLKPLAGVYRSLKKQIEKNITFTNLNLDILNSRLKRRKYLVDVL 122
Db :||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7 IDNILSKIENEVEVLKPLAGVYRSLKKQLENNVMTFNVNVKIDILNSRFNKNFKNVL 66
QY 123 ESDLMOFKHISNEYIIEDSFLLNSEQKNTLLKSKYIKESVENDIKFAQEGISYYEKV 182
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 67 ESDLIPYKDLTSSNVYVVDYKPLNKRKDKFLSSNYIKDSIDTDFINFANDVLYGYKIL 126
QY 183 LAKYKDDLESIKKVIKEKEKFPSPPTTPPSPAKTDEQKESKFLPFLTNITLYNNLV 242
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 127 SEKYSKDLDSIKYI:-----NDKQGENEKYLPFLNNITLYKTVN 166
QY 243 NKIDYLYNLKAKINDCNVKEDEAHVITKLSDLKAIDDKIDLFKNPYDFEAIKKLINDD 302
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 167 DKIDLFIHLEAKVLNVTYKESNVKIKELNVLTKTIQDKLADFKNNNFVGTADLSTDY 226
QY 303 TKKDMGLKLLSTGLV-ONFPNTIISKLIBCKFQDMLNISQHCVKKQCPENSGCFRHLDE 361
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 227 NNINLLTKFLSGMVFENIAKTVLNLNLDGNLQGMINTISOHCVKKQCPQNSGCFRHLDE 286
QY 362 REECKLLNYKQEGDKCVENPNPTNENNGGDAADATCTEEDSGSRKKITCECTKPDYSY 421
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 287 REECKLLNYKQEGDKCVENPNPTNENNGGDAADAKCTEEDSGSGNGKKITCECTKPDYSY 346
QY 422 PLFDGIFCS 430
Db 347 PLFDGIFCS 355
RESULT 4
```

```
AAAY05832
ID AAY05832 standard; Protein; 355 AA.
XX
AC AAY05832;
XX
DT 02-AUG-1999 (first entry)
XX
DE Merozoite surface protein MSP-1-42.
XX
KW MSP-1; merozoite surface protein; malaria; vaccine;
KW protein engineering; protein expression; codon usage;
KW transgenic animal.
XX
OS Plasmodium falciparum.
XX
PN WO9920766-A2.
XX
PD 29-APR-1999.
XX
PF 20-OCT-1998; 98WO-US222225.
XX
PR 15-MAY-1998; 98US-0085649.
PR 20-OCT-1997; 97US-0062592.
XX
PA (GENZ ) GENZYME TRANSGENICS CORP.
XX
PI Chen LH, Meade H;
XX
DR WPI: 1999-302742/25.
DR N-PSDB; AAX25586.
XX
CC New modified recombinant nucleic acid sequences useful for producing
CC malarial DNA vaccine
CC Disclosure; Fig 1; 43pp; English.
XX
CC The present sequence represents a 42 kDa C-terminal portion of
CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC important target for the development of a vaccine against
CC Plasmodium falciparum. The nucleic acid sequence encoding MSP-1-42
CC has been modified (see AAX25586) compared to the native sequence (see
CC AAX25587) such that 306 nucleotide positions have been replaced to
CC lower the AT content (from 76 to 49.7%) and to eliminate 10 mRNA
CC instability motifs while maintaining the same protein amino acid
CC sequence. These alterations allow MSP-1-42 to be expressed in
CC mammalian cell culture and in transgenic mice. The invention
CC provides modified recombinant nucleic acid sequences and methods for
CC increasing the mRNA levels and protein expression of proteins that
CC are difficult to express in cell culture systems, mammalian cell
CC culture systems or in transgenic animals. The preferred difficult
CC protein candidates for expression are those derived from lower
CC organisms such as parasites, bacteria and viruses that have DNA
CC coding sequences of high AT content or which have mRNA instability
CC motifs or rare codons relative to the recombinant expression system
CC to be used. The invention allows expression of MSP-1 protein in
CC the milk of transgenic animals, and also provides a DNA vaccine
CC comprising a vector containing the altered MSP-1-42 sequence.
XX
SQ Sequence 355 AA;
Query Match 51.1%; Score 1161.5; DB 20; Length 355;
Best Local Similarity 59.1%; Pred. No. 8.8e-75;
Matches 218; Conservative 54; Mismatches 76; Indels 21; Gaps 2;
QY 63 MNILSGFENEVDVYLKPLAGVYRSLKKQIEKNITFTNLNLDILNSRLKRRKYLVDVL 122
Db :||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7 IDNILSKIENEVEVLKPLAGVYRSLKKQLENNVMTFNVNVKIDILNSRFNKNFKNVL 66
QY 123 ESDLMOFKHISNEYIIEDSFLLNSEQKNTLLKSKYIKESVENDIKFAQEGISYYEKV 182
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 67 ESDLIPYKDLTSSNVYVVDYKPLNKRKDKFLSSNYIKDSIDTDFINFANDVLYGYKIL 126
QY 183 LAKYKDDLESIKKVIKEKEKFPSPPTTPPSPAKTDEQKESKFLPFLTNITLYNNLV 242
```

```
Db 127 SEKYSOLDSTKKYI-----NDKQGENEKYLPFLNNIETLYKTVN 166
QY 243 NKIDYILINKAKTNDNCNEKDEAHVITKILSLDKATDDKIDLFKNPYDFEAIKKLINDD 302
Db 167 DKIDLFVHLEAKVNTYKESNVEVAKIKELNYLKTQDKLADFKKNNFVGIADLSTDY 226
QY 303 TKKDMGLKLLSTGLV-QNFPNTIISKIEGKFQDMLNISQHCYKVKQCPENSGCFRHLDE 361
Db 227 NHNLLTKFLSTGMVFENLAKTVLSNLLDGNLQGLMNLISQHCYKVKQCPQNSGCFRHLDE 286
QY 362 REECKCLLNYKQEGDKCVENPNPTCNENNGCGDADATCTEEDSGSSRKKITCECTKPDYSY 421
Db 287 REECKCLLNYKQEGDKCVENPNPTCNENNGCGDADAKCTEEDSGSNGKKITCECTKPDYSY 346
QY 422 PLFDGIFCS 430
Db 347 PLFDGIFCS 355

RESULT 5
AAAY09373
ID AAY09373 standard; Protein; 361 AA.
XX
AC AAY09373;
XX
DT 31-AUG-1999 (first entry)
DE Merozoite surface protein MSP-1-42.
XX
KW MSP-1; merozoite surface protein; malaria; vaccine;
KW protein engineering; protein expression; codon usage;
KW transgenic animal.
XX
OS Plasmodium falciparum.
XX
PN WO9920774-A2.
XX
PD 29-APR-1999.
XX
PF 20-OCT-1998; 98WO-US22226.
XX
PR 15-MAY-1998; 98US-0085649.
PR 20-OCT-1997; 97US-0062592.
XX
PA (GENZ ) GENZYME TRANSGENICS CORP.
XX
PI Chen LH, Meade H;
XX
DR WPI; 1999-288313/24.
DR P-PSDB; AAX56009.
XX
PT Modified malarial protein for use in anti-malarial vaccines
XX
PS Example; Fig 2; 35pp; English.
XX
CC This present sequence comprises a 42 kDa C-terminal portion of
CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC important target for the development of a vaccine against
CC Plasmodium falciparum. The C-terminal end of the sequence is
CC modified to include a 6xHis tag. A nucleic acid (see AAX56008)
CC encoding MSP-1-42 has been modified according to a method
CC of the invention in order to improve expression in mammalian cells
CC and in transgenic animals by reducing the AT content and removing
CC mRNA instability motifs. The invention allows expression of
CC MSP-1-42 in the milk of transgenic animals, and also provides a DNA
CC vaccine comprising a vector containing the altered MSP-1-42 nucleic
CC acid.
XX
SQ Sequence 361 AA;
Query Match 51.1%; Score 1161.5; DB 20; Length 361;
Best Local Similarity 59.1%; Pred. No. 9e-75;
```

```
Matches 218; Conservative 54; Mismatches 76; Indels 21; Gaps 2;
QY 63 MDNLTSGFENEYDIYLLKPLAGVYRSLLKQIEKNIFTNLNDILNSRLKRRKYFELDVL 122
Db 7 IDNLSIKIENEYEVLYLKLPLAGVYRSLLKQLENNVMTFNVKDILNSRKNKRFNRVL 66
QY 123 ESOLMOPKHSSNRYIIDSFKLNSQKNTLLSKYIKESVENDIKFAQEGISYIEKV 182
Db 67 ESDLIPYKDLTSSNYYVVKDPYKFLNKEKROKFLSSYNIKDSIDTDINFANDVLGYKIL 126
QY 183 LAKYKDDLESITKKVIEKEEFSPPTTPPSAKTDEOKKESKFLPLNFIETLYNNLV 242
Db 127 SEKYSOLDSTKKYI-----NDKQGENEKYLPFLNNIETLYKTVN 166
QY 243 NKIDYILINKAKTNDNCNEKDEAHVITKILSLDKATDDKIDLFKNPYDFEAIKKLINDD 302
Db 167 DKIDLFVHLEAKVNTYKESNVEVAKIKELNYLKTQDKLADFKKNNFVGIADLSTDY 226
QY 303 TKKDMGLKLLSTGLV-QNFPNTIISKIEGKFQDMLNISQHCYKVKQCPENSGCFRHLDE 361
Db 227 NHNLLTKFLSTGMVFENLAKTVLSNLLDGNLQGLMNLISQHCYKVKQCPQNSGCFRHLDE 286
QY 362 REECKCLLNYKQEGDKCVENPNPTCNENNGCGDADATCTEEDSGSSRKKITCECTKPDYSY 421
Db 287 REECKCLLNYKQEGDKCVENPNPTCNENNGCGDADAKCTEEDSGSNGKKITCECTKPDYSY 346
QY 422 PLFDGIFCS 430
Db 347 PLFDGIFCS 355

RESULT 6
AAAY05833
ID AAY05833 standard; Protein; 361 AA.
XX
AC AAY05833;
XX
DT 02-AUG-1999 (first entry)
DE Merozoite surface protein MSP-1-42.
XX
KW MSP-1; merozoite surface protein; malaria; vaccine;
KW protein engineering; protein expression; codon usage;
KW transgenic animal.
XX
OS Plasmodium falciparum.
XX
PN WO9920766-A2.
XX
PD 29-APR-1999.
XX
PF 20-OCT-1998; 98WO-US22225.
XX
PR 15-MAY-1998; 98US-0085649.
PR 20-OCT-1997; 97US-0062592.
XX
PA (GENZ ) GENZYME TRANSGENICS CORP.
XX
PI Chen LH, Meade H;
XX
DR WPI; 1999-302742/25.
DR N-PSDB; AAX25587.
XX
PT New modified recombinant nucleic acid sequences useful for producing
PT malarial DNA vaccine
XX
PS Disclosure; Fig 2; 43pp; English.
XX
CC This present sequence comprises a 42 kDa C-terminal portion of
CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC important target for the development of a vaccine against
CC Plasmodium falciparum. The C-terminal end of the sequence is
CC modified to include a 6xHis tag. Nucleic acids (see AAX25586 and
```



```

XX 02-AUG-1999 (first entry)
XX Modified merozoite surface protein MSP-1-42.
DE MSP-1; merozoite surface protein; malaria; vaccine;
KW protein engineering; protein expression; codon usage;
KW transgenic animal; mutant.
XX Plasmodium falciparum.
OS Synthetic.
XX Key Location/Qualifiers
FH Peptide 1..15
FT /note= "beta-casein signal peptide"
FT Peptide 371..376
FT /note= "6xHis tag"
FT Misc-difference 197
FT /note= "Asn in native MSP-1-42 (N181Q mutation)"
FT Misc-difference 278
FT /note= "Asn in native MSP-1-42 (N262Q mutation)"
XX WO9920766-A2.
XX 29-APR-1999.
XX 20-OCT-1998; 98WO-US22225.
XX 15-MAY-1998; 98US-0085649.
XX 20-OCT-1997; 97US-0062592.
XX (GENZ ) GENZYME TRANSGENICS CORP.
XX Chen LH, Meade H;
PI WPI; 1999-302742/25.
XX N-PSDB; AAX25593.
XX New modified recombinant nucleic acid sequences useful for producing
XX malarial DNA vaccine
XX Example; Fig 11; 43pp; English.
XX The present sequence represents a 42 kDa C-terminal portion of
XX malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
XX important target for the development of a vaccine against
XX Plasmodium falciparum. The sequence has been modified to include
XX an N-terminal beta-casein signal peptide and an N-terminal 6xHis
XX tag. In addition, N181Q and N262Q mutations have been introduced to
XX eliminate 2 N-glycosylation sites. These modifications allow the
XX MSP-1-42 protein to be expressed in the milk of transgenic mice.
XX The invention generally relates to modified recombinant nucleic
XX acid sequences and methods for increasing the mRNA levels and
XX protein expression of proteins that are difficult to express in
XX cell culture systems, mammalian cell culture systems or in
XX transgenic animals. The preferred difficult protein candidates for
XX expression are those derived from lower organisms such as parasites,
XX bacteria and viruses that have DNA coding sequences of high AT
XX content or which have mRNA instability motifs or rare codons
XX relative to the recombinant expression system to be used.
XX Sequence 376 AA;
XX Query Match 50.7%; Score 1154.5; DB 20; Length 376;
XX Best Local Similarity 58.8%; Pred. No. 3e-74;
XX Matches 217; Conservative 53; Mismatches 78; Indels 21; Gaps 2;
XX 63 MDNLSGFENEYDVIYKPLAGVYSLKKQIEKNFTFNLDNLNLSRLKRRYFLDVL 122
XX :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
XX 22 IDNLSKIENEYEVLYLPLAGVYSLKKQLENNVMTNVNKKDILNSRENKRENFKNVL 81
XX :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
XX 123 RSDLMQFHISSEYIIEEDSKFLNLSKQNTLLKSKYIKESVENDIKFAEGISYIEYK 182
XX :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

```

```

Db 82 ESDLIPYKDLTSSNYYVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYKIL 141
Qy 183 LAKYKDDLESITKKVKEEKEKFPSPPTTSPAKTDEQKESKFLPLTNLTLYNNLV 242
Db 142 SERYSKDLDSIKYI-----NDKQGENEKYLPFLNNIETLYKTVN 181
Qy 243 NKIDDYLINLAKINDCNVEKDEAHVKITKLSDLKATDDKIDLFKNPYDFEAIKKLINDD 302
Db 182 DKIDLFIHLEAKVLQTYEKSNNVEVKIKELNYLKTQDKLADFKNNNEVGIADLSTDY 241
Qy 303 TKKDMGLKLLSTGLV-QNFPNTIISKLEIFQDMLNISQHCVKKQCPENSGCFRHLDE 361
Db 242 NHNNLLTKFLSTGMVFENLAKTVLSNLLDGNLQMLQISQHCYKQCPQNSGCFRHLDE 301
Qy 362 REECKCLLNKQEGDKCVENPNTCENNGGCDADATCTEEDSGSSRKKITCTECTKPDYS 421
Db 302 REECKCLLNKQEGDKCVENPNTCENNGGCDADAKCTEEDSGSNGKKITCTECTKPDYS 361
Qy 422 PLFDGIFCS 430
Db 362 PLFDGIFCS 370
XX RESULT 9
XX AAP50777
XX ID AAP50777 standard; Protein: 1654 AA.
XX AC AAP50777;
XX XX
XX DT 30-SEP-1991 (first entry)
XX XX
XX DE Sequence of the P195 protein of Plasmodium falciparum.
XX KW Malaria vaccine; epitope; antigen; immunogen.
XX OS Plasmodium falciparum.
XX PN EP154454-A.
XX XX
XX PD 11-SEP-1985.
XX XX
XX PF 21-FEB-1985; 85EP-0301173.
XX PR 26-SEP-1984; 84GB-0024340.
XX PR 22-FEB-1984; 84GB-0004692.
XX PR 21-FEB-1985; 85GB-0004429.
XX XX
XX PA (WELL ) WELLCOME FOUNDATION LTD.
XX PI Holder A, Sandhu J, Odink K, Lockyer M, Riveros-Moreno V;
XX DR WPI; 1985-224845/37.
XX N-PSDB; AAN50530.
XX XX
XX PT Cloned DNA sequence encoding plasmodium falciparum protein -
XX PT useful for expressing the protein for use in vaccines against
XX PT malaria
XX XX
XX PS Claim 6; Fig 1; 51pp; English.
XX CC The sequence encoding the P195 protein of Plasmodium falciparum
XX CC (AAN50530) and a peptide comprising at least one of its epitopes
XX CC (see AAP50777) are claimed. Also claimed is a vaccine for inducing
XX CC immunity to malaria comprising the novel peptide or P195 or a
XX CC peptide comprising at least one epitope when derived from the new
XX CC DNA sequence, together with a carrier.
XX SQ Sequence 1654 AA;
XX Query Match 44.3%; Score 1007.5; DB 6; Length 1654;
XX Best Local Similarity 50.9%; Pred. No. 6.6e-63;
XX Matches 202; Conservative 65; Mismatches 105; Indels 25; Gaps 5;

```

QY 37 GTDDDDKADAGISIEGRCTMAISVTMONILSGFENEYDVIYKPLAGVYRSLLKKOIEKN 96  
DB 1261 GESEED--YDDLQGVVTVGEAVTTSV-IDNLSKIENEYEVLYKPLAGVYRSLLKKQLENN 1317  
QY 97 IITFNLNLDILNSRLKKRYFLDVLSEDLMOFKHISSEYIIEFSKLLNSEQKNTLLK 156  
DB 1318 VMTFNVNVDILNSRFKNKFNKLVLESLIPYKOLTSYVYKDPYKFLNKKRDKFLS 1377  
QY 157 SYKIKESVENDIKFAQEGISYVEKVLAKYKDDLESIKKVIKEEKEKFPSPPTTPPSA 216  
DB 1378 SYNVIKSDTDINFAVNDVLYKYLSEKYSKDLDSIKKYI----- 1418  
QY 217 KTDQKESKFLPFLNIETLYNNLVNKIDDLINLAKINDCNVEKDEAHVKITKLSDL 276  
DB 1419 -NDKOGENEKYLPLFLNLTLYKTVDKIDLFVHLEAKVLNTEKSNVEVKIKELNVL 1477  
QY 277 KATDDKIDLFKNPYDPEAKKLINDTKDMLGKLLSTGLV-QNFPNTIISKLECKF-Q 334  
DB 1478 KTIQDKLADFKNKNNNFVGIADLSTDYNNHNNLTGFLSTGVMFENLKSLSNLDWKLAR 1537  
QY 335 DMLNISQHCVRKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGGCD 394  
DB 1538 YVKHFTTPMRKKTMIQSSGCFRHLDERECKCLLNYKQEGSKCVENSNPTCNENNGGCD 1597  
QY 395 ADATCTEEDSGSRKKITCECTKPDSPYPLFDGIFCSS 431  
DB 1598 ADAKCTEEDSGSGKKITCECTKPCDYPPLSMVIFCSS 1634  
RESULT 10  
AAB37610  
ID AAB37610 standard; Protein; 262 AA.  
AC AAB37610;  
XX  
DT 27-FEB-2001 (first entry)  
XX  
DE Merozoite surface protein-133.  
XX  
KW Merozoite surface protein; protazoacide; vaccine; malaria.  
XX  
OS Plasmodium falciparum.  
XX  
PN WO200063245-A2.  
XX  
PD 26-OCT-2000.  
XX  
PF 20-APR-2000; 2000WO-CB01558.  
XX  
PR 20-APR-1999; 99GB-0009072.  
PR 13-MAY-1999; 99US-0311817.  
PR 25-MAY-1999; 99CA-2271451.  
XX  
PA (MEDI-) MEDICAL RES COUNCIL.  
PI Holder A, Birdsall B, Feeney J, Morgan W, Syed S, Uthaipibull C;  
XX  
DR WPI; 2001-015762/02.  
DR N-PSDB; AAC68978.  
XX  
PT Novel variants of the C-terminal fragment of Plasmodium merozoite surface protein-1, useful as vaccines for treating or preventing malaria.  
XX  
PS Example 5; Fig 15; 126pp; English.  
XX  
CC The present invention relates to non-natural variants of a C-terminal fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The non-natural variants have reduced affinity for at least 1 antibody capable of blocking a second antibody that inhibits the proteolytic cleavage of Plasmodium MSP-1.4.2, and has the same affinity for at least one third antibody that inhibits the proteolytic cleavage of Plasmodium MSP-1.4.2, compared to natural MSP-1.1.9. The non-natural variants of the

CC present invention are useful for immunising a mammal against malaria, and can be used to treat malaria. The present sequence is MSP-133 protein.  
XX  
SQ Sequence 262 AA;  
Query Match 28.2%; Score 640.5; DB 22; Length 262;  
Best Local Similarity 46.7%; Pred. No. 8.7e-38;  
Matches 129; Conservative 51; Mismatches 75; Indels 21; Gaps 2;  
QY 63 MDNLSGFENEYDVIYKPLAGVYRSLLKKOIEKNITFNLNLDILNSRLKKRYFLDVL 122  
DB 7 IDNLSKIENEYEVLYKPLAGVYRSLLKQLENNVMTFNVNVDILNSRFKNKFNKVL 66  
QY 123 ESDLMOFKHISSEYIIEFSKLLNSEQKNTLLKSYKIKESVENDIKFAQEGISYVEK 182  
DB 67 ESDLIPYKOLTSYVYKDPYKFLNKKRDKFLSSYNIKSDTDINFAVNDVLYKYL 126  
QY 183 LAKYKDDLESIKKVIKEEKEKFPSPPTTPPSAKTDEQKESKFLPFLNIETLYNNLV 242  
DB 127 SEKYKSDLSIKKYI-----NDKOGENEKYLPLFLNLTLYKTVD 166  
QY 243 NKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLKATDDKIDLFKNPYDPEAKKLIND 302  
DB 167 DKIDLFVHLEAKVLQYTEKSNVEVKIKELNLTGFLSTGVMFENLKSLSNLDWKLAR 226  
QY 303 TKKDMGLKLLSTGLV-QNFPNTIISKLECKFQDML 337  
DB 227 NHNNLTGFLSTGVMFENLAKTVLSNLDGLOGML 262  
RESULT 11  
AAB36103  
ID AAB36103 standard; Protein; 116 AA.  
XX  
AC AAB36103;  
XX  
DT 25-MAR-1998 (first entry)  
XX  
DE PfMSP1(p19)A protein sequence.  
XX  
KW Plasmodium vivax; merozoite surface protein; MSP1; p19;  
KW Plasmodium falciparum; malaria; vaccine; immunity; epitope.  
XX  
OS Plasmodium falciparum.  
OS Synthetic.  
FH Key Location/Qualifiers  
FT Region 1..95 /note= "amino acids derived from P. falciparum MSP1 p19 fragment"  
FT Region 96..116 /note= "glycosylphosphatidylinositol anchoring sequence"  
PN WO9730158-A2.  
XX  
PD 21-AUG-1997.  
XX  
PF 14-FEB-1997; 97WO-FR00290.  
XX  
PR 14-FEB-1996; 96FR-0001822.  
XX  
PA (INSP ) INST PASTEUR.  
PA (UUNY ) UNIV NEW YORK STATE.  
XX  
PI Barnwell JW, Longacre-Andre S, Mendis K, Nato F;  
PI Roth C;  
XX  
DR WPI; 1997-425033/39.  
DR N-PSDB; AAT94550.  
XX  
PT Recombinant protein containing the merozoite surface protein-1 p19 fragment - useful in anti-malarial vaccines, diagnosis and protein purification  
PT

XX PS Disclosure; Fig 1B; 85pp; French.

XX CC This is the amino acid sequence of a recombinant protein comprising

CC amino acids 1613-1705 of the Plasmodium falciparum merozoite surface

CC protein 1 (MSPl) 19 kD C-terminal fragment (p19), linked to a

CC glycosylphosphatidylinositol membrane anchoring sequence. p19 is the

CC C-terminal fragment of the 42 kD MSPl from Plasmodium species.

CC The recombinant protein can be used for the production of anti-malarial

CC vaccines, where the p19 fragment provides a high level of protective

CC immunity since it includes epitopes not presented in the p42 fragment.

XX Sequence 116 AA;

XX Query Match 23.2%; Score 527; DB 18; Length 116;

XX Best Local Similarity 96.8%; Pred. No. 3.7e-30;

XX Matches 91; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 338 NISQHCYKVKQCPENSGCFRHLDERECKCLLNKYQEGDKCVENPNTCNENGGCDADA 397

DB 3 NISQHCYKVKQCPENSGCFRHLDERECKCLLNKYQEGDKCVENPNTCNENGGCDADA 62

QY 398 TCTEEDSGSRKKITCECTKPDSPFLFDGIFCSS 431

DB 63 KCTEEDSGSGKKITCECTKPDSPFLFDGIFCSS 96

RESULT 12

AAW22592

ID AAW22592 standard; Protein; 116 AA.

XX AC AAW22592;

XX DT 25-MAR-1998 (first entry)

XX DE PfMSPl(p19)A protein sequence.

XX KW Plasmodium vivax; merozoite surface protein; MSPl; p19;

XX KW Plasmodium falciparum; malaria; vaccine; immunity; epitope.

XX OS Plasmodium falciparum.

XX OS Synthetic.

XX FH Key Location/Qualifiers

FT Region 1..95 /note= "amino acids derived from P. falciparum MSPl p19 fragment"

FT Region 96..116 /note= "glycosylphosphatidylinositol anchoring sequence"

XX WO9730159-A2.

XX 21-AUG-1997.

XX 14-FEB-1997; 97WO-FR00291.

XX 14-FEB-1996; 96FR-0001821.

XX (INSP ) INST PASTEUR.

XX PA (UYNV ) UNIV NEW YORK STATE.

XX PI Barnwell JW, Longacre-Andre S, Mendis K, Nato F;

XX Roth C;

XX WPI; 1997-425034/39.

XX P-PSDB; AAW22592.

XX Recombinant protein containing Plasmodium merozoite surface

XX protein-1 p42 fragment - useful in antimalarial vaccines, also new

XX antibodies for diagnosis and protein purification

XX PS Disclosure; Fig 1B; 85pp; French.

CC This is the amino acid sequence of a recombinant protein comprising

CC amino acids 1613-1705 of the Plasmodium falciparum merozoite surface

CC protein 1 (MSPl) 19 kD C-terminal fragment (p19), linked to a

CC glycosylphosphatidylinositol membrane anchoring sequence. p19 is the

CC C-terminal fragment of the 42 kD MSPl from Plasmodium species.

CC The recombinant protein can be used for the production of anti-malarial

CC vaccines, where the p19 fragment provides a high level of protective

CC immunity since it includes epitopes not presented in the p42 fragment.

XX Sequence 116 AA;

XX Query Match 23.2%; Score 527; DB 18; Length 116;

XX Best Local Similarity 96.8%; Pred. No. 3.7e-30;

XX Matches 91; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 338 NISQHCYKVKQCPENSGCFRHLDERECKCLLNKYQEGDKCVENPNTCNENGGCDADA 397

DB 3 NISQHCYKVKQCPENSGCFRHLDERECKCLLNKYQEGDKCVENPNTCNENGGCDADA 62

QY 398 TCTEEDSGSRKKITCECTKPDSPFLFDGIFCSS 431

DB 63 KCTEEDSGSGKKITCECTKPDSPFLFDGIFCSS 96

RESULT 13

AAW22593

ID AAW22593 standard; Protein; 127 AA.

XX AC AAW22593;

XX DT 25-MAR-1998 (first entry)

XX DE PfMSPl(p19)S protein sequence.

XX KW Chimeric; Plasmodium vivax; merozoite surface protein; MSPl; p19;

XX KW Plasmodium falciparum; malaria; vaccine; immunity; epitope.

XX OS Chimeric - Plasmodium vivax.

XX OS Chimeric - Plasmodium falciparum.

XX FH Key Location/Qualifiers

FT Peptide 1..19 /note= "signal peptide"

FT Protein 20..127 /note= "mature protein"

FT Region 1..32 /note= "derived from P. vivax MSPl"

FT Region 33..34 /note= "encoded by restriction enzyme sequence used to create the chimeric sequence"

FT Region 35..127 /note= "derived from P. falciparum C-terminal p19 fragment of MSPl"

XX WO9730159-A2.

XX 21-AUG-1997.

XX 14-FEB-1997; 97WO-FR00291.

XX 14-FEB-1996; 96FR-0001821.

XX (INSP ) INST PASTEUR.

XX PA (UYNV ) UNIV NEW YORK STATE.

XX PI Barnwell JW, Longacre-Andre S, Mendis K, Nato F;

XX Roth C;

XX WPI; 1997-425034/39.

XX P-PSDB; AAW22592.

XX Recombinant protein containing Plasmodium merozoite surface

XX protein-1 p42 fragment - useful in antimalarial vaccines, also new

PT antibodies for diagnosis and protein purification  
 XX Disclosure; Fig 1C; 85pp; French.  
 XX This is the amino acid sequence of a chimeric protein comprising amino  
 CC acids 1-32 of the Plasmodium vivax merozoite surface protein 1 (MSP1)  
 CC linked to amino acids Asn1613-Ser1705 which correspond to the 19 kD  
 CC C-terminal fragment (p19) of MSP-1 from Plasmodium falciparum.  
 CC p19 is the C-terminal fragment of the 42 kD MSP1 from Plasmodium species.  
 CC The recombinant protein can be used for the production of anti-malarial  
 CC vaccines, where the p19 fragment provides a high level of protective  
 CC immunity since it includes epitopes not presented in the p42 fragment.  
 XX Sequence 127 AA;  
 SQ Query Match 23.2%; Score 527; DB 18; Length 127;  
 Best Local Similarity 94.8%; Pred. No. 4.1e-30;  
 Matches 91; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 QY 335 DMLNISQHCVKKQCPENSGCFRHLDERECKCLLNKQEGDKCVENPNTCNENNGGCD 394  
 DB 32 DEFNISQHCVKKQCPENSGCFRHLDERECKCLLNKQEGDKCVENPNTCNENNGGCD 91  
 QY 395 ADATCTEEDSGSRKKITCECTKPDSPYPLFDGIFCS 430  
 DB 92 ADAKCTEEDSGSGRKKITCECTKPDSPYPLFDGIFCS 127  
 RESULT 14  
 AAW36102  
 ID AAW36102 standard; Protein; 127 AA.  
 XX AAW36102;  
 XX 25-MAR-1998 (first entry)  
 XX pfMSP1(p19)S protein sequence.  
 XX Chimeric; Plasmodium vivax; merozoite surface protein; MSP1; p19;  
 KW Plasmodium falciparum; malaria; vaccine; immunity; epitope.  
 XX Chimeric - Plasmodium vivax.  
 OS Chimeric - Plasmodium falciparum.  
 XX Key Location/Qualifiers  
 FT Peptide 1..19  
 FT /note= "signal peptide"  
 FT Protein 20..127  
 FT /note= "mature protein"  
 FT Region 1..32  
 FT /note= "derived from P. vivax MSP1"  
 FT Region 33..34  
 FT /note= "encoded by restriction enzyme sequence used to  
 FT create the chimeric sequence"  
 FT Region 35..127  
 FT /note= "derived from P. falciparum C-terminal p19  
 FT fragment of MSP1"  
 XX WO9730158-A2.  
 XX 21-AUG-1997.  
 XX 14-FEB-1997; 97WO-FR00290.  
 XX 14-FEB-1996; 96FR-0001822.  
 XX (INSP ) INST PASTEUR.  
 PA (UUNY ) UNIV NEW YORK STATE.  
 XX Barwell JW, Longacre-Andre S, Mendis K, Nato F;  
 PI Roth C;  
 XX WPI; 1997-425033/39.

DR N-PSDB; AAT94549.  
 XX Recombinant protein containing the merozoite surface protein-1 p19  
 PT fragment - useful in anti-malarial vaccines, diagnosis and protein  
 PT purification  
 XX Disclosure; Fig 1C; 85pp; French.  
 XX This is the amino acid sequence of a chimeric protein comprising amino  
 CC acids 1-32 of the Plasmodium vivax merozoite surface protein 1 (MSP1)  
 CC linked to amino acids Asn1613-Ser1705 which correspond to the 19 kD  
 CC C-terminal fragment (p19) of MSP-1 from Plasmodium falciparum.  
 CC p19 is the C-terminal fragment of the 42 kD MSP1 from Plasmodium species.  
 CC The recombinant protein can be used for the production of anti-malarial  
 CC vaccines, where the p19 fragment provides a high level of protective  
 CC immunity since it includes epitopes not presented in the p42 fragment.  
 XX Sequence 127 AA;  
 SQ Query Match 23.2%; Score 527; DB 18; Length 127;  
 Best Local Similarity 94.8%; Pred. No. 4.1e-30;  
 Matches 91; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 QY 335 DMLNISQHCVKKQCPENSGCFRHLDERECKCLLNKQEGDKCVENPNTCNENNGGCD 394  
 DB 32 DEFNISQHCVKKQCPENSGCFRHLDERECKCLLNKQEGDKCVENPNTCNENNGGCD 91  
 QY 395 ADATCTEEDSGSRKKITCECTKPDSPYPLFDGIFCS 430  
 DB 92 ADAKCTEEDSGSGRKKITCECTKPDSPYPLFDGIFCS 127  
 RESULT 15  
 AAB37609  
 ID AAB37609 standard; Protein; 108 AA.  
 XX AAB37609;  
 XX 27-FEB-2001 (first entry)  
 XX Merozoite surface protein-119.  
 XX Merozoite surface protein; protozoacide; vaccine; malaria.  
 XX Plasmodium falciparum.  
 XX WO2000063245-A2.  
 XX 26-OCT-2000.  
 XX 20-APR-2000; 2000WO-GB01558.  
 XX 20-APR-1999; 99GB-0009072.  
 XX 13-MAY-1999; 99US-0311817.  
 XX 25-MAY-1999; 99CA-2271451.  
 XX (MEDI-) MEDICAL RES COUNCIL.  
 XX Holder A, Birdsell B, Feeney J, Morgan W, Syed S, Uthaipibull C;  
 XX WPI; 2001-015762/02.  
 DR N-PSDB; AAC68977.  
 XX Novel variants of the C-terminal fragment of Plasmodium merozoite  
 PT surface protein-1, useful as vaccines for treating or preventing  
 PT malaria -  
 XX Example 5; Fig 15; 126pp; English.  
 XX The present invention relates to non-natural variants of a C-terminal  
 CC fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The  
 CC non-natural variants have reduced affinity for at least 1 antibody  
 CC capable of blocking a second antibody that inhibits the proteolytic

CC cleavage of Plasmodium MSP-1.4\_2, and has the same affinity for at least  
 CC one third antibody that inhibits the proteolytic cleavage of Plasmodium  
 CC MSP-1.4\_2, compared to natural MSP-1.1.9. The non-natural variants of the  
 CC present invention are useful for immunising a mammal against malaria, and  
 CC can be used to treat malaria. The present sequence is MSP-119 protein.

XX SQ Sequence 108 AA;

Query Match 23.1%; Score 525; DB 22; Length 108;

Best Local Similarity 89.3%; Pred. No. 4.6e-30;

Matches 92; Conservative 4; Mismatches 3; Indels 4; Gaps 1;

Qy 329 IEGFQDMLNTSQHCVKQCPNSGCFRHLDERECKLLNYKQEGDKCVENPNTCNE 388

Db 8 IEGRH---NTAQHCVKQCPNSGCFRHLDERECKLLNYKQEGDKCVENPNTCNE 63

Qy 389 NNGGCDADATCEEDSGSRKKTCTCKPDSYPLFDGIFCSS 431

Db 64 NNGGCDADAKTEEDSGSGNGKKTCTCKPDSYPLFDGIFCSS 106

RESULT 16

AAB37608

ID AAB37608 standard; protein; 96 AA.

XX AAB37608;

DT 27-FEB-2001 (first entry)

DE Merozoite surface protein-1.

KW Merozoite surface protein; protazoacide; vaccine; malaria.

OS Plasmodium falciparum.

PN WO200063245-A2.

XX 26-OCT-2000.

XX 20-APR-2000; 2000WO-GB01558.

PR 20-APR-1999; 99GB-0009072.

PR 13-MAY-1999; 99US-0311817.

PR 25-MAY-1999; 99CA-2271451.

XX (MEDI-) MEDICAL RES COUNCIL.

PI Holder A, Birdsall B, Feeney J, Morgan W, Syed S, Uthaiipibull C;

DR WPT; 2001-015762/02.

PT Novel variants of the C-terminal fragment of Plasmodium merozoite  
 PT surface protein-1, useful as vaccines for treating or preventing  
 PT malaria -

PS Example 2; Page 48; 126pp; English.

XX The present invention relates to non-natural variants of a C-terminal  
 CC fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The  
 CC non-natural variants have reduced affinity for at least 1 antibody  
 CC capable of blocking a second antibody that inhibits the proteolytic  
 CC cleavage of Plasmodium MSP-1.4\_2, and has the same affinity for at least  
 CC one third antibody that inhibits the proteolytic cleavage of Plasmodium  
 CC MSP-1.4\_2, compared to natural MSP-1.1.9. The present sequence is the  
 CC wild-type MSP-1 protein. This sequence was used to generate the variants  
 CC of the present invention. The non-natural variants of the present  
 CC invention are useful for immunising a mammal against malaria, and can be  
 CC used to treat malaria.

XX SQ Sequence 96 AA;

Query Match

Best Local Similarity 23.0%; Score 524; DB 22; Length 96;

Matches 95.7%; Pred. No. 4.7e-30;

Matches 90; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 338 NISOHCVKKQCPNSGCFRHLDERECKLLNYKQEGDKCVENPNTCENNGCCDADA 397

Db 1 NISOHCVKKQCPNSGCFRHLDERECKLLNYKQEGDKCVENPNTCENNGCCDADA 60

Qy 398 TCTEEDSGSRKKTCTCKPDSYPLFDGIFCSS 431

Db 61 KCTEEDSGSGNGKKTCTCKPDSYPLFDGIFCSS 94

RESULT 17

AAU00669

ID AAU00669 standard; Protein; 206 AA.

XX AAU00669;

DT 07-SEP-2001 (first entry)

DE Plasmodium vivax merozoite surface protein C-terminal region.

KW Merozoite surface protein; malaria; blood; serum; diagnosis; vaccine;  
 KW antigen-antibody composite; Enzyme Linked Immunosorbent Assay; ELISA;  
 KW MSP.

XX Plasmodium vivax.

XX WO200136587-A2.

XX 25-MAY-2001.

XX 15-NOV-2000; 2000WO-KR01302.

XX 15-NOV-1999; 99KR-0050616.

PR 25-APR-2000; 2000KR-0022041.

PR 20-MAY-2000; 2000KR-0027305.

XX (HUMA-) HUMANBIO CO LTD.

XX Park H;

XX WPT; 2001-343809/36.

DR N-PSDB; AAS00655.

XX New gene encoding merozoite surface protein of Plasmodium vivax, useful  
 for producing protein for diagnosis of malaria and for vaccination -

XX Claim 5; Page 24; 24pp; English.

XX The sequence represents a Plasmodium vivax merozoite surface protein  
 CC (MSP) C-terminal region. The C-terminal region of the merozoite surface  
 CC protein has a strong antigenicity in malarial diseases. For diagnosis of  
 CC malaria, recombinant proteins with enhanced antigenicity, obtained by  
 CC addition of fusion proteins to surface protein C-terminal regions, can be  
 CC reacted with serum or blood of a Plasmodium infected patient.  
 CC Antigen-antibody composites will be formed, and these are detected by  
 CC Enzyme Linked Immunosorbent Assay (ELISA). The recombinant antigens  
 CC provide a quick and reliable diagnosis of malaria, with good sensitivity  
 CC and selectivity.

XX SQ Sequence 206 AA;

Query Match 17.8%; Score 406; DB 22; Length 206;

Best Local Similarity 41.0%; Pred. No. 3.4e-21;

Matches 80; Conservative 40; Mismatches 63; Indels 12; Gaps 4;

Qy 238 YNNLVNKIDDDYLINLAKKINDCNVEKDEAHVKITKLSDLKAIDDKIDLKKNPYDFEALKK 297

Db 2 YESLVSVNTYTDNLKKVINNCQLEKAEITVKKLDQDYNKMKDEKLEBK-----KSEKK 56

Qy 298 LINDDTKKDMLGKLTGLV-QNEPNTIISKLEGGKFDMLNLSHQHCVKQCPNSGCF 356

Db 57 --NEVKSGLLEKLMKSLIKENESKEITLSQLLVQVLTQTTMSSEHTCIDTNVPDNRACY 114









KW Multivalent protein; immune response; Plasmodium vivax; parasite;  
 KW protozoacide; vaccine; malaria; recombinant; ViVac2.

OS Synthetic.

OS Plasmodium vivax.

PN WO200155181-A2.

XX 02-AUG-2001.

XX 29-JAN-2001; 2001WO-US02937.

XX 31-JAN-2000; 2000US-0179213.

XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.

XX Lal AA, Xiao L, Zhou Z;

XX WPI; 2001-514557/56.

DR N-PSDB; AAH47055.

XX New recombinant multivalent protein comprising antigenic determinants  
 PT derived from more than one stage in a life cycle of Plasmodium vivax,  
 PT useful as a vaccine for treating, preventing and reducing malarial  
 PT infection

XX Claim 5; Page 48-55; 59pp; English.

XX The invention relates to recombinant multivalent proteins (I) that  
 CC stimulate an immune response to Plasmodium vivax. (I) comprises antigenic  
 CC determinants, fragments or conservative substitutions, derived from more  
 CC than one stage in a life cycle of a Plasmodium vivax parasite. (I) is  
 CC useful as a vaccine for stimulating an immune response, specifically a  
 CC protective immune response that confers increased resistance to infection  
 CC by Plasmodium parasites, such as P. vivax. (I) is especially useful in  
 CC the treatment, prevention and reduction of malarial infection, as  
 CC research or diagnostic reagents for the detection of Plasmodium species  
 CC in a biological sample, and for conferring immunity against multiple  
 CC stages of the malarial parasite. The antibodies produced are useful for  
 CC the detection or measurement of antigenic epitopes derived from one or  
 CC more stages in a life cycle of a parasite, particularly P. vivax. The  
 CC vaccine comprising the recombinant proteins, is cost-effective, health-  
 CC promoting intervention for controlling, preventing or treating the  
 CC incidence of malaria. The present sequence represents the amino acid  
 CC sequence of the recombinant protein ViVac2p, a multivalent and  
 CC multistage vaccine against P. vivax.

XX Sequence 2028 AA;

Query Match 12.5%; Score 283.5; DB 22; Length 2028;

Best Local Similarity 24.2%; Pred. No. 3.9e-11;

Matches 86; Conservative 52; Mismatches 108; Indels 109; Gaps 11;

QY 88 SLKKQIEKNITFNILNLIINSRLKKRYFLDVLSDLMQFKHISSEYIIEFSKLLN 147

Db 1771 SVKRLKGN-FWICKLVAVN-----IEQIVKRWIREGRDYVSE----LPT 1813

QY 148 SEQNTLLKSYKIKESVENDIKF-----AQEGISYKVKLAKYKD--DLESTKK 195

Db 1814 EVQK-----LKEKCDGKINTDKKVKVPCQACKSVDQWTRKKNQWDVLSNKF 1864

QY 196 VIKKEKEKFPSPPTPPSPAKTDEQKESKFLPFLTNIEFLYNNLVNKIDDYILNLIKAK 255

Db 1865 ISVKNKRVQTAGIVTPYDILKQELDEFN-----EVAFENEINKRDGAVIELCV- 1913

QY 256 INDCNVEKDEAHVITKLSLDLKAIDDKIDLPKNPYDEPAIKLLINDTTKKDMLKLLSTG 315.

Db 1914 ---CSVEE-----AKNTQEVVTNVDNAKSPWTMS----- 1941

QY 316 LVQNFPTIISKIEGKFDQMLNLSOHCVKVKCPENSGCFRHLDERECKCLLNKOEK 375

Db 1942 -----SEHTCIDTNVDPNACRYILDGTEWRCLLTFRKEG 1977

QY 376 DKCVENPNPTCNENNGCCDADATCTEEDSGSRKKITCECTKPDSPYFLDGIKCS 430  
 Db 1978 GKCVPASNVTKDNNNGCAPEAECKMTDS---NKIVCKTKEGSEPPFBEVFC 2028

RESULT 24

AAR41355

ID AAR41355 standard; peptide; 49 AA.

XX AAR41355;

XX 04-MAR-1994 (first entry)

XX MSPLEGFIB EGF1-like domain variant.

XX Epidermal growth factor 1; merozoite surface protein 1; malaria;

KW vaccine.

XX Plasmodium yoelii.

XX Key Location/Qualifiers

FT Cleavage-site 1

FT /note= "introduced to facilitate cleavage

FT from recombinant protein"

XX WO9317107-A.

XX 02-SEP-1993.

XX 22-FEB-1993; 93WO-GB00367.

XX 22-FEB-1992; 92GB-0003821.

XX (MEDI-) MEDICAL RES COUNCIL.

XX Blackman MJ, Chappel JA, Holder AA;

XX WPI; 1993-288413/36.

XX Allelic variants of epidermal growth factor 1- or 2-like domains - of  
 PT merozoite surface protein 1, produced recombinantly for malaria  
 PT vaccines

XX Claim 1; Fig 1b; 35pp; English.

XX The sequence is that of an allelic variant of a merozoite surface  
 CC protein-1 epidermal growth factor (EGF) 1 like domain. It may be  
 CC used alone or as part of a fusion protein of EGF-1-like and  
 CC EGF-2-like domains in vaccines against malaria. When expressed  
 CC recombinantly it is produced in a form indistinguishable from  
 CC that in the native protein.

XX Sequence 49 AA;

Query Match 12.4%; Score 282; DB 14; Length 49;

Best Local Similarity 98.0%; Pred. No. 3.5e-13;

Matches 48; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 337 LNTISOHCVKKQCPENSGCFRHLDERECKCLLNKOEKQKCVENPNPT 385

Db 1 MNTISOHCVKKQCPENSGCFRHLDERECKCLLNKOEKQKCVENPNPT 49

RESULT 25

AAB47487

ID AAB47487 standard; peptide; 108 AA.

XX AAB47487;

XX 15-NOV-2001 (first entry)

XX PV200C polypeptide.

KW PV200C; merozoite surface protein; Plasmodium vivax; malaria;  
 KW malaria-specific antibody; blood; antigen; protozoa; immunoassay.  
 OS Plasmodium vivax.

XX WO200161032-A1.

XX 23-AUG-2001.

XX 15-FEB-2001; 2001WO-KR00229.

XX 17-FEB-2000; 2000KR-0007648.

XX 17-FEB-2000; 2000KR-0007649.

XX 10-MAR-2000; 2000KR-0007650.

XX 08-AUG-2000; 2000KR-0012172.

XX 08-AUG-2000; 2000KR-0045806.

XX (GLDS ) LG CHEM LTD.

XX Lim K, Shon M, Yoo S, Lee S, Oh J, Lee S, Kim H;

XX WPI; 2001-565352/63.

XX Detecting malaria-specific antibodies in blood using antigens of

XX malarial protozoa, useful for diagnosing malaria with long latent

XX periods and low blood Protozoa levels -

XX Claim 11; Page 86-87; 89pp; English.

XX This sequence is the polypeptide PV200C which is derived from the

XX C-terminal portion of the merozoite surface protein of P. vivax. This

XX polypeptide may be used in an immunoassay for malaria, which involves

XX detecting malaria-specific antibodies in blood by using antigens of

XX the malarial protozoa. The immunoassay has high specificity/sensitivity

XX and may be used to diagnose types of malaria in which the latent

XX period is long and the number of protozoa found in the blood is low.

XX PV200C has high specificity to the antibodies and high purity. The

XX PV200C surface protein has low pseudo-positive signals.

XX Sequence 108 AA;

XX Query Match 12.3%; Score 280; DB 22; Length 108;

XX Best Local Similarity 45.5%; Pred. No. 1.4e-12;

XX Matches 51; Conservative 19; Mismatches 38; Indels 4; Gaps 1;

XX QY 319 NFPTNTIISKIEGKFQDMLNISQHCVKQCPENSGCFRHLDERECKLLNYKQEGDKC 378

XX Db 1 NESKEILSOLLNVQTLTWSSHTCIDTNVPDNCACYRLDGTETWRCLLTFKEGGKC 60

XX QY 379 VNPNTCNENNGCDADATCTEEDSGSRKKKITCCTKPDSPYLPFDGIFCS 430

XX Db 61 VPASNVYTKDNNNGCAPEACKMTDS----NKIVCKCTKESGSEPLFEGVFC 108

XX RESULT 26

XX AAR41354

XX ID AAR41354 standard; peptide; 49 AA.

XX XX

XX AC AAR41354;

XX XX

XX DT 04-MAR-1994 (first entry)

XX DE MSPLEGFLA EGF1-like domain variant.

XX XX

XX KW Epidermal growth factor 1; merozoite surface protein 1; malaria;

XX KW vaccine.

XX OS Plasmodium yoelii.

XX XX

XX FH Key Location/Qualifiers

XX FT Cleavage-site 1

XX FT /note= "introduced to facilitate cleavage

XX FT from recombinant protein"

XX FT

XX WO9317107-A.  
 XX 02-SEP-1993.  
 XX 22-FEB-1993; 93WO-GB00367.  
 XX 22-FEB-1992; 92GB-0003821.  
 XX (MEDI-) MEDICAL RES COUNCIL.  
 XX Blackman MJ, Chappel JA, Holder AA;  
 XX WPI; 1993-288413/36.  
 XX Allelic variants of epidermal growth factor 1- or 2-like domains - of  
 XX merozoite surface protein 1, produced recombinantly for malaria  
 XX vaccines  
 XX Claim 1; Fig 1a; 35pp; English.  
 XX The sequence is that of an allelic variant of a merozoite surface  
 XX protein-1 epidermal growth factor (EGF) 1 like domain. It may be  
 XX used alone or as part of a fusion protein of EGF-1-like and  
 XX EGF-2-like domains in vaccines against malaria. When expressed  
 XX recombinantly it is produced in a form indistinguishable from  
 XX that in the native protein.  
 XX Sequence 49 AA;  
 XX Query Match 12.3%; Score 279; DB 14; Length 49;  
 XX Best Local Similarity 95.9%; Pred. No. 5.8e-13;  
 XX Matches 47; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 XX QY 337 LNIHQVCVKQCPENSGCFRHLDERECKLLNYKQEGDKCVENPNPT 385  
 XX Db 1 MNISQHCVKQCPQNSGCFRHLDERECKLLNYKQEGDKCVENPNPT 49  
 XX RESULT 27  
 XX AEEL3033  
 XX ID AEEL3033 standard; Protein; 289 AA.  
 XX AC AEEL3033;  
 XX 28-JAN-2002 (first entry)  
 XX Helicobacter pylori His-HOP38(-11) polypeptide.  
 XX Polypeptide delivery system; immune stimulating complex; ISCOM;  
 XX cholesterol; saponin; phospholipid; medicament; vaccine; therapy;  
 XX Helicobacter pylori infection; HOP38; anti-bacterial.  
 XX Helicobacter pylori.  
 XX WO200176625-A1.  
 XX 18-OCT-2001.  
 XX 09-APR-2001; 2001WO-SE00799.  
 XX 12-APR-2000; 2000GB-0008879.  
 XX (ASTR ) ASTRAZENECA AB.  
 XX Shapero A, Sanyal G;  
 XX WPI; 2001-663016/76.  
 XX N-PSDB; AAD21371.  
 XX Producing a polypeptide delivery system useful in a vaccine to treat  
 XX infection by mixing together the polypeptide, cholesterol, saponin, and  
 XX a phospholipid in presence of a nonionic detergent and a second

PT detergent  
 PS Disclosure: Page 40; 43pp; English.  
 XX  
 CC The invention relates to producing a polypeptide delivery system  
 CC comprising an immune stimulating complex (ISCOM) coupled to a  
 CC polypeptide. The method involves mixing the polypeptide, cholesterol,  
 CC a saponin and a phospholipid in the presence of a nonionic detergent  
 CC and a second detergent to form a solution and removing the detergent  
 CC from the mixture to form the ISCOM. The polypeptide delivery system  
 CC is used as a medicament for the manufacture of a vaccine for  
 CC administration to mammalian patients, to treat and prevent  
 CC Helicobacter pylori infection. The present sequence is His-HOP38(-11),  
 CC a Helicobacter pylori polypeptide used in the polypeptide delivery  
 CC system.  
 XX  
 SQ Sequence 289 AA;  
 Query Match 12.2%; Score 278.5; DB 22; Length 289;  
 Best Local Similarity 96.4%; Pred. No. 6.6e-12;  
 Matches 53; Conservative 0; Mismatches 1; Indels 1; Gaps 1;  
 OY 1 MHHHHHSSGLVPRGSGMKETAARFERQHMDSPDLGTDGDDKAMADIGSIEGRG 55  
 DB 1 MHHHHHSSGLVPRGSGMKETAARFERQHMDSPDLGTDGDDKAMADIGS-EGDG 54  
 RESULT 28  
 AAB47732  
 ID AAB47732 standard; Protein: 289 AA.  
 XX  
 AC AAB47732;  
 XX  
 DT 30-JAN-2002 (first entry)  
 XX  
 DE HOP38 #5.  
 XX  
 KW H. pylori; HOP38; polypeptide delivery system; polar head group;  
 KW immune stimulating complex; ISCOM; antigenic fragment; saponin;  
 KW cholesterol; phospholipid; detergent.  
 OS Helicobacter pylori.  
 PN WO200176623-A1.  
 XX  
 PD 18-OCT-2001.  
 XX  
 PF 09-APR-2001; 2001WO-SE00800.  
 XX  
 PR 12-APR-2000; 2000GB-0008877.  
 XX  
 PA (ASTR ) ASTRAZENECA AB..  
 XX  
 PI Shapero A, Sanyal G;  
 XX  
 DR WPI: 2002-025884/03.  
 DR N-PSDB: AAH43782.  
 XX  
 CC Production of a polypeptide delivery system useful as a medicament  
 CC comprises mixing together the polypeptide, cholesterol, saponin, and a  
 CC phospholipid with a polar head group, in the presence of a detergent -  
 PS Disclosure: Page 45; 48pp; English.  
 XX  
 CC The sequences given in AAB47728-32 are H. pylori HOP38 proteins  
 CC which were used in the method of the invention. The method of the  
 CC invention is a process for production of a polypeptide delivery system  
 CC comprising an immune stimulating complex (ISCOM) coupled to a  
 CC polypeptide of H. pylori or its antigenic fragment. The method comprises:  
 CC mixing the polypeptide, cholesterol, a saponin, and a phospholipid  
 CC having a polar head group; and removing the detergent from the mixture  
 CC to form an ISCOM. The method has a broad applicability to polypeptides,  
 CC including polypeptides that are unsuited to prior art processes.

XX  
 SQ Sequence 289 AA;  
 Query Match 12.2%; Score 278.5; DB 23; Length 289;  
 Best Local Similarity 96.4%; Pred. No. 6.6e-12;  
 Matches 53; Conservative 0; Mismatches 1; Indels 1; Gaps 1;  
 OY 1 MHHHHHSSGLVPRGSGMKETAARFERQHMDSPDLGTDGDDKAMADIGSIEGRG 55  
 DB 1 MHHHHHSSGLVPRGSGMKETAARFERQHMDSPDLGTDGDDKAMADIGS-EGDG 54  
 RESULT 29  
 AAM48247  
 ID AAM48247 standard; Protein: 824 AA.  
 XX  
 AC AAM48247;  
 XX  
 DT 21-MAR-2002 (first entry)  
 XX  
 DE Thioredoxin/deoxyxylulose 5-phosphate synthase fusion protein.  
 XX  
 KW DXPS; deoxyxylulose 5-phosphate synthase; pyruvate; thioredoxin;  
 KW glyceraldehyde 3-phosphate; plant growth modulator;  
 KW microbial growth modulator; enzyme.  
 XX  
 OS Chimeric - Arabidopsis thaliana.  
 OS Chimeric - Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 1..165  
 FT /note= "Thioredoxin protein"  
 FT Region 166..824  
 FT /note= "Truncated DXPS protein"  
 XX  
 PN US6326164-B1.  
 XX  
 PD 04-DEC-2001.  
 XX  
 PF 27-JUL-2000; 2000US-0626589.  
 XX  
 PR 27-JUL-2000; 2000US-0626589.  
 XX  
 PA (PARA-) PARADIGM GENETICS INC.  
 XX  
 PI Rice JW, Kloti AS, Crawford JM, Lanning B, Stewart SJ;  
 XX  
 DR WPI: 2002-121106/16.  
 DR N-PSDB: ABA95618.  
 XX  
 PT Assay for determining deoxyxylulose 5-phosphate synthase activity by  
 PT measuring pyruvate depletion, useful for screening compounds that  
 PT inhibit or enhance this activity which is useful for modulating plant  
 PT and microbial growth -  
 XX  
 PS Claim 1; Columns 19-24; 26pp; English.  
 CC  
 CC The present sequence is a fusion protein comprising a truncated  
 CC deoxyxylulose 5-phosphate synthase (DXPS) from Arabidopsis thaliana and  
 CC thioredoxin (trxA). The N-terminal 58 amino acids of the DXPS protein  
 CC were removed to generate the truncated protein. The truncated DXPS protein  
 CC possesses DXPS activity. The full-length DXPS is given in AAM48245. The  
 CC present invention relates to a method for determining DXPS-activity. The  
 CC method comprises contacting pyruvate, and optionally glyceraldehyde  
 CC 3-phosphate with DXPS and then determining the concentration of remaining  
 CC pyruvate and/or glyceraldehyde 3-phosphate. The method is useful for  
 CC screening for inhibitors and enhancers of DXPS activity which will have  
 CC use as modulators of plant and microbial growth.

XX  
 SQ Sequence 824 AA;  
 Query Match 12.2%; Score 278.5; DB 23; Length 824;  
 Best Local Similarity 37.4%; Pred. No. 2.7e-11;

Matches 76; Conservative 25; Mismatches 49; Indels 53; Gaps 7;

QY 1 MHHHHSSGLVPRGSGMKETAARERQHMDSPLDGTDDDDKAMADIGSIEGRGTMAIS 60  
|||||  
Db 116 MHHHHSSGLVPRGSGMKETAARERQHMDSPLDGTDDDDKAMADIGSAS----- 167  
QY 61 VTMDNLGFGFENEYDVIYKPLA--GVYSLKKQIEKNIFTNLDNLNLSKLRKYF 118  
||| | | : : : : :  
Db 168 -----LAERGEYS-NRPPTPLDITINYPH-MKNLSVKELKOL 204  
QY 119 LDVLESDLMQFK-----HISSEYIIEDSKLLNSEQNTLLSKYIKESVENDIKFA 172  
||| | | : : : : :  
Db 205 SDELRSDVI-FNVSKTGGHGLSGLVVE-----LTVALHYIFNTPODKILWD 250  
QY 173 QEGISYERKVLAKYKDDLESIKK 195  
||| | | : : : : :  
Db 251 VGHQSYPHKILTGRRCKMPTMRQ 273

RESULT 30  
AAR41356  
ID AAR41356 standard; peptide; 54 AA.  
XX  
AC AAR41356;  
XX  
DT 04-MAR-1994 (first entry)  
XX  
DE MSPLEGF2A EGF2-like domain variant.  
XX  
KW Epidermal growth factor 1; merozoite surface protein 1; malaria;  
KW vaccine.  
XX  
OS Plasmodium yoelii.  
XX  
FH Key Location/Qualifiers  
FT Cleavage-site 1  
FT /note= "introduced to facilitate cleavage  
FT from recombinant protein"

XX  
PN W09317107-A.  
XX  
XX  
PD 02-SEP-1993.  
XX  
PF 22-FEB-1993; 93WO-CB00367.  
XX  
PR 22-FEB-1992; 92GB-0003821.  
XX  
PA (MEDI-) MEDICAL RES COUNCIL.  
XX  
PI Blackman MJ, Chappel JA, Holder AA;  
XX  
XX WPI; 1993-288413/36.  
XX  
PT Allelic variants of epidermal growth factor 1- or 2-like domains - of  
PT merozoite surface protein 1, produced recombinantly for malaria  
PT vaccines  
XX  
PS Claim 2; Fig 2a; 35pp; English.  
XX  
CC The sequence is that of an allelic variant of a merozoite surface  
CC protein-1 epidermal growth factor (EGF) 2-like domain. It may be  
CC used alone or as part of a fusion protein of EGF-1-like and  
CC EGF-2-like domains in vaccines against malaria. When expressed  
CC recombinantly it is produced in a form indistinguishable from  
CC that in the native protein.  
XX  
SQ Sequence 54 AA;  
Query Match 12.2%; Score 278; DB 14; Length 54;  
Best Local Similarity 94.1%; Pred. No. 7.7e-13;  
Matches 48; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 381 NPNTCNENNGCDATCTEEDSGSRKKITCECTKPDSPYPLFDGIFCSS 431

Db 2 NPNTCNENNGCDADAKCTEEDSGNGKKITCECTKPDSPYPLFDGIFCSS 52  
|||||  
RESULT 31  
AAB11532  
ID AAB11532 standard; Protein; 692 AA.  
XX  
AC AAB11532;  
XX  
DT 12-OCT-2000 (first entry)  
XX  
DE SEN virus protein fragment SEQ ID NO: 104.  
XX  
KW SEN virus; SENV; gastrointestinal tract disorder; inflammatory disease;  
KW proliferative disorder; hepatopathy; hepatitis; viral infection;  
KW vaccination; gene therapy.  
XX  
OS Hepatitis virus.  
XX  
PN W0200028039-A2.  
XX  
PD 18-MAY-2000.  
XX  
PF 09-NOV-1999; 99WO-EP08566.  
XX  
PR 10-NOV-1998; 98IT-MI02437.  
PR 30-APR-1999; 99IT-MI00923.  
PR 14-MAY-1999; 99EP-0830298.  
PR 16-JUL-1999; 99EP-0113932.  
XX  
PA (DIAS-) DIASORIN SRL.  
XX  
PI Primi D, Fiordalisi G, Mantero GL, Mattioli S, Sottini A;  
PI Bonelli F, Vaglini L, Olivero P, Dal Corso A, Bonelli M;  
XX  
DR WPI; 2000-376551/32.  
XX  
PT Nucleic acids representing the genome of the SEN virus (SENV) and  
PT encoded proteins, useful for treatment of hepatopathies, inflammatory  
PT diseases and proliferative disorders such as cancer -  
XX  
PS Example 16; Page 78; 392pp; English.  
XX  
CC The present invention is concerned with the sequence of the genome of the  
CC SEN virus (SENV), and the proteins encoded by it. SENV is thought to be  
CC the cause of hepatopathies which are not linked to the presence of the  
CC hepatitis A, B and E viruses in man. The genome and proteins of this  
CC virus can be used in gene therapy and vaccination against the virus,  
CC which also causes disorders of the gastrointestinal tract, including  
CC Crohn's disease and lupus erythematosus, inflammatory diseases, and  
CC proliferative disorders such as cancer.  
XX  
SQ Sequence 692 AA;  
Query Match 12.2%; Score 277.5; DB 21; Length 692;  
Best Local Similarity 46.9%; Pred. No. 2.5e-11;  
Matches 67; Conservative 12; Mismatches 37; Indels 27; Gaps 4;

QY 1 MHHHHSSGLVPRGSGMKETAARERQHMDSPLDGTDDDDKAMADIGSIEGRGTMAIS 60  
|||||  
Db 1 MHHHHSSGLVPRGSGMKETAARERQHMDSPLDGTDDDDKAMADIGSMN----- 52  
QY 61 VTMDNLGFGFENEYDVIYKPLAGVYVRSLLKQIEKNIFTN-LNLNDILNLSRLKKRYFL 119  
: : : : :  
Db 53 -----YAMHCEDSTPQPEFGG-----DMSTITFSLLYLDQHERHLNRWTFPN 96  
QY 120 DVLESDLMQFKHISSEYIIEDS 142  
||| | | : : : : :  
Db 97 DQL--DLVRYKHTFRFKYRSKDT 117

RESULT 32

AAY85150  
 ID AAY85150 standard; protein; 467 AA.  
 XX  
 AC AAY85150;  
 XX  
 DT 23-JUN-2000 (first entry)  
 XX  
 DE Mouse secreted protein acidic and rich in cysteine (SPARC).  
 XX  
 KW SPARC; secreted protein acidic and rich in cysteine; pharmaceutical;  
 KW drug; research reagent; mouse.  
 XX  
 OS Mus musculus.  
 XX  
 PN JP3012931-B1.  
 XX  
 PD 28-FEB-2000.  
 XX  
 PF 26-FEB-1999; 99JP-0049826.  
 XX  
 PR 26-FEB-1999; 99JP-0049826.  
 XX  
 PA (AGEN ) KOGYO GIJUTSUINCHO.  
 XX  
 DR WPI: 2000-259560/23.  
 DR N-PSDB; AAA09770.  
 XX  
 PT New secretion protein acidic and rich in cysteine (SPARC) for drugs and  
 PT research reagents consists of specific amino acid sequence -  
 XX  
 PS Claim 1; Fig 1; 17pp; Japanese.  
 XX  
 CC This sequence represents the amino acid sequence of a mouse protein,  
 CC having the physiological activity of SPARC (secreted protein acidic and  
 CC rich in cysteine). The protein is used for the preparation of drugs,  
 CC research reagents and pharmaceuticals. The SPARC protein is prepared  
 CC efficiently and inexpensively in large quantities. The protein can be  
 CC used in pharmaceuticals as it does not lose its physiological activity.  
 XX  
 SQ Sequence 467 AA;  
 Query Match 12.1%; Score 275; DB 21; Length 467;  
 Best Local Similarity 98.0%; Pred. No. 2.2e-11;  
 Matches 50; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MHHHHHSSGLVPRGSGMKETAARFQHMDSPLDGLTDDDDKAMADIGSI 51  
 DB 116 MHHHHHSSGLVPRGSGMKETAARFQHMDSPLDGLTDDDDKAMADIGSM 166  
 RESULT 33  
 AAY79067  
 ID AAY79067 standard; protein; 467 AA.  
 XX  
 AC AAY79067;  
 XX  
 DT 20-JUN-2000 (first entry)  
 XX  
 DE Murine secreted protein acidic and rich in cysteine amino acid sequence.  
 XX  
 KW Secreted protein acidic and rich in cysteine; SPARC; mouse; nerve cell;  
 KW neurocyte cell adhesion; cell protrusion retraction; cell shrinkage;  
 KW nervous system disease; epilepsy; arteriosclerosis; wound healing.  
 XX  
 OS Mus sp.  
 XX  
 PN JP3012930-B1.  
 XX  
 PD 28-FEB-2000.  
 XX  
 PF 26-FEB-1999; 99JP-0049708.  
 XX  
 PR 26-FEB-1999; 99JP-0049708.

XX  
 PA (AGEN ) KOGYO GIJUTSUINCHO.  
 XX  
 DR WPI: 2000-306484/27.  
 DR N-PSDB; AAZ98759.  
 XX  
 PT Drug composition for suppressing neurocyte cell adhesion, generating  
 PT cell migration and promoting shrinkage retraction of the cell contains  
 PT specific amino acid sequence -  
 XX  
 PS Claim 1; Fig 4; 21pp; Japanese.  
 XX  
 CC This sequence represents the amino acid sequence of the murine protein  
 CC referred to as SPARC (secreted protein acidic and rich in cysteine). The  
 CC protein is used as the active ingredient in a drug composition for  
 CC suppressing neurocyte cell adhesion, generating cell migration and  
 CC promoting a nerve protrusion shrinkage reaction. The composition can be  
 CC used in wound healing and also to treat diseases of the nervous system,  
 CC arteriosclerosis, and epilepsy.  
 XX  
 SQ Sequence 467 AA;  
 Query Match 12.1%; Score 275; DB 21; Length 467;  
 Best Local Similarity 98.0%; Pred. No. 2.2e-11;  
 Matches 50; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MHHHHHSSGLVPRGSGMKETAARFQHMDSPLDGLTDDDDKAMADIGSI 51  
 DB 116 MHHHHHSSGLVPRGSGMKETAARFQHMDSPLDGLTDDDDKAMADIGSM 166  
 RESULT 34  
 AAB26110  
 ID AAB26110 standard; protein; 76 AA.  
 XX  
 AC AAB26110;  
 XX  
 DT 30-JAN-2001 (first entry)  
 XX  
 DE H. contortus clone 65e vector pET30a cloning junction #2.  
 XX  
 KW Nematode; parasite; helminth; sheep; goat; stomach; vaccine.  
 XX  
 OS Synthetic.  
 XX  
 PN WO2000056763-A1.  
 XX  
 PD 28-SEP-2000.  
 XX  
 PF 16-MAR-2000; 2000WO-AU00210.  
 XX  
 PR 18-MAR-1999; 99AU-0009297.  
 XX  
 PA (NOVS ) NOVARTIS AG.  
 PA (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.  
 XX  
 PI Savin KW, Cook VR, Chen Y, Sexton JL, Apos E, Wilson LR;  
 PI Griffiths TM, Newton SE;  
 XX  
 DR WPI: 2000-594573/56.  
 DR N-PSDB; AAA94064.  
 XX  
 PT New Haemonchus contortus polypeptide for inducing a protective effect  
 PT against a helminth by controlling helminth infection, growth, viability  
 PT and/or egg fecundity and for ameliorating the symptoms of helminth  
 PT infection -  
 XX  
 PS Disclosure; Fig 8; 94pp; English.  
 XX  
 CC The present sequence is the translated sequence found at the cloning  
 CC junction in the pET30a vector of nematode Haemonchus contortus  
 CC clone 65e. This clone encodes a novel protein. This organism is a  
 CC parasite found in the stomach of its host (sheep and goats in



CC particular). The protein can be used in a vaccine against other  
 CC helminths, as well as Haemonchus contortus, including trematodes,  
 CC cestodes, nematodes and acanthocephala. These are all capable of causing  
 CC severe illness in their hosts, which include sheep, pigs, goats, cattle,  
 CC horses, donkeys, dogs, cats, guinea pigs and cage-birds, along with  
 CC humans. Antibodies to the protein can be used to diagnose infection.  
 CC Note: This sequence is stated as being the same as that shown in SEQ ID  
 CC NO: 6 of the specification (see AAB26109). However, this sequence is  
 CC shorter than the one shown here.

XX  
 SQ Sequence 76 AA;

Query Match 12.0%; Score 274; DB 21; Length 76;  
 Best Local Similarity 100.0%; Pred. No. 2.4e-12;

Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHSSGLVPRGSGMKETAATAAKFERQHMDSPDLGTDGDDDKAMADIGS 50  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1 MHHHHHSSGLVPRGSGMKETAATAAKFERQHMDSPDLGTDGDDDKAMADIGS 50

RESULT 35

AAB85253  
 ID AAB85253 standard; Protein; 167 AA.

XX  
 AC AAB85253;

XX 07-SEP-2001 (first entry)

DE Thioresoxin functional fragment.

XX Thioresoxin; porphobilinogen synthase; T-PPS; PPS; enzyme; herbicide;  
 KW delta-aminolevulinic acid; plant growth inhibitor.

XX Lycopersicon esculentum.

XX WO200146446-A1.

XX 28-JUN-2001.

XX 19-DEC-2000; 2000WO-US34584.

XX 22-DEC-1999; 99US-0171785.

XX (PARA-) PARADIGM GENETICS INC.

XX Crawford JM, Rice J, Sevala V, Stewart S;

XX WPI; 2001-418081/44.

DR N-PSDB; AAH22802.

XX Novel plant thioresoxin-porphobilinogen synthase or porphobilinogen  
 PT synthase polypeptides, useful for identifying compounds for use as  
 PT herbicides by inhibiting enzymatic activity of the polypeptides -

XX Example 1; Page 19; 25pp; English.

XX The invention provides novel DNA sequences encoding enzymes such as plant  
 CC thioresoxin-porphobilinogen synthase (T-PPS) and plant porphobilinogen  
 CC synthase (PPS). Methods of determining the enzymatic activity of T-PPS or  
 CC PPS or its functional fragment are provided that involves contacting  
 CC delta-aminolevulinic acid with the protein or its functional fragment and  
 CC measuring the amount of porphobilinogen formed from it. Compounds which  
 CC can modify the enzymatic activity of T-PPS or PPS can also be identified  
 CC similarly, which are useful for inhibiting plant growth by inhibiting  
 CC enzymatic activity of T-PPS or PPS or its functional fragment. The  
 CC compounds thus identified are useful as herbicides. The present sequence  
 CC represents the thioresoxin functional fragment.

XX Sequence 167 AA;

Query Match 12.0%; Score 274; DB 22; Length 167;  
 Best Local Similarity 100.0%; Pred. No. 6.7e-12;

Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHSSGLVPRGSGMKETAATAAKFERQHMDSPDLGTDGDDDKAMADIGS 50  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 116 MHHHHHSSGLVPRGSGMKETAATAAKFERQHMDSPDLGTDGDDDKAMADIGS 165

RESULT 36

AAB85254  
 ID AAB85254 standard; Protein; 167 AA.

XX  
 AC AAB85254;

XX 07-SEP-2001 (first entry)

XX Thioresoxin functional fragment.

XX Thioresoxin; porphobilinogen synthase; T-PPS; PPS; enzyme; herbicide;  
 KW delta-aminolevulinic acid; plant growth inhibitor.

XX Lycopersicon esculentum.

XX WO200146446-A1.

XX 28-JUN-2001.

XX 19-DEC-2000; 2000WO-US34584.

XX 22-DEC-1999; 99US-0171785.

XX (PARA-) PARADIGM GENETICS INC.

XX Crawford JM, Rice J, Sevala V, Stewart S;

XX WPI; 2001-418081/44.

DR N-PSDB; AAH22802.

XX Novel plant thioresoxin-porphobilinogen synthase or porphobilinogen  
 PT synthase polypeptides, useful for identifying compounds for use as  
 PT herbicides by inhibiting enzymatic activity of the polypeptides -

XX Example 1; Page 20; 25pp; English.

XX The invention provides novel DNA sequences encoding enzymes such as plant  
 CC thioresoxin-porphobilinogen synthase (T-PPS) and plant porphobilinogen  
 CC synthase (PPS). Methods of determining the enzymatic activity of T-PPS or  
 CC PPS or its functional fragment are provided that involves contacting  
 CC delta-aminolevulinic acid with the protein or its functional fragment and  
 CC measuring the amount of porphobilinogen formed from it. Compounds which  
 CC can modify the enzymatic activity of T-PPS or PPS can also be identified  
 CC similarly, which are useful for inhibiting plant growth by inhibiting  
 CC enzymatic activity of T-PPS or PPS or its functional fragment. The  
 CC compounds thus identified are useful as herbicides. The present sequence  
 CC represents the thioresoxin functional fragment.

XX Sequence 167 AA;

Query Match 12.0%; Score 274; DB 22; Length 167;  
 Best Local Similarity 100.0%; Pred. No. 6.7e-12;

Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHSSGLVPRGSGMKETAATAAKFERQHMDSPDLGTDGDDDKAMADIGS 50  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 116 MHHHHHSSGLVPRGSGMKETAATAAKFERQHMDSPDLGTDGDDDKAMADIGS 165

RESULT 37

AAB85251  
 ID AAB85251 standard; Protein; 551 AA.

XX  
 AC AAB85251;

XX 07-SEP-2001 (first entry)







GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 19, 2003, 14:17:05 : Search time 218 Seconds  
(without alignments)  
4452.345 Million cell updates/sec

Title: US-10-057-531A-2

Perfect score: 2275

Sequence: 1 MHHHHHSSGLVPRGSMKE.....TCECTKPDSPYLFDGIFCSS 431

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cgn2.1/USPTO.spool/US10057531/runat\_12052003.091030.21654/app\_query.fasta\_1.503  
-DB=N\_Geneseq\_101002 -QFMT=fastp -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=40  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10057531 -RCGN\_1\_1\_200\_@runat\_12052003.091030.21654 -NCPU=6 -ICPU=3  
-NO\_XLPT=NO\_MMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCAPOP=6 -FCAPEXT=7  
-YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_101002.\*

1: /SID2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1970	86.6	1128	22	DNA encoding major
2	1931.5	84.9	1181	16	Plasmodium falciparum
3	1175	51.6	1897	18	Chimeric MSA-1 ant
4	1175	51.6	1950	18	Chimeric MSA-1 ant
5	1175	51.6	4940	19	P. falciparum mod
6	1172	51.5	1896	18	Chimeric MSA-1 ant
7	1161.5	51.1	1065	20	Merzoite surface
8	1161.5	51.1	1065	20	Merzoite surface
9	1161.5	51.1	1077	22	Merzoite surface
10	1161.5	51.1	1088	20	Merzoite surface
11	1161.5	51.1	1088	20	Merzoite surface
12	1161	51.0	4940	19	P. falciparum gpl9
13	1154.5	50.7	1140	20	Merzoite surface
14	1154.5	50.7	1140	20	Merzoite surface
15	1066	46.9	5760	6	Sequence encoding
16	640.5	28.2	786	22	Merzoite surface
17	527	23.2	354	18	PFMSPI(pl9)A codin
18	527	23.2	354	18	PFMSPI(pl9)A codin
19	527	23.2	387	18	PFMSPI(pl9)S codin
20	527	23.2	387	18	PFMSPI(pl9)S codin
21	525	23.1	333	22	Merzoite surface
22	521	22.9	309	24	Plasmodium falcipa
23	521	22.9	3147	24	Plasmodium falcipa
24	521	22.9	3147	24	Human C3d3 DNA-pf
25	500	22.0	309	24	Plasmodium falcipa
26	500	22.0	3147	24	Mutant PFMSPI.19-h
27	500	22.0	3147	24	Human C3d3-pfMSPI.
28	406	17.8	618	22	Plasmodium vivax m
29	284.5	12.5	5438	22	Synthetic gene viv
30	284.5	12.5	6101	22	Synthetic gene viv
31	283.5	12.5	5643	22	PFMSPI.19-murine c
32	282.5	12.4	3252	24	Helicobacter pylor
33	278.5	12.2	870	22	HOP38 coding sequ
34	278.5	12.2	870	24	Thioredoxin/deoxyx
35	278.5	12.2	2472	24	SEN virus genome f
36	277.5	12.2	2079	21	Secreted protein a
37	275	12.1	7375	21	Murine SPARC nucle
38	275	12.1	7375	21	H. contortus clone
39	274	12.0	241	21	Clone 6 DNA relate
40	274	12.0	428	22	Thioredoxin functi
41	274	12.0	501	22	Clone 7 DNA relate
42	274	12.0	589	22	Plant thioredoxin-
43	274	12.0	1653	22	Plasmid pGR3 encod
44	274	12.0	6553	24	Plasmid pGR7 encod
45	274	12.0	6618	24	

ALIGNMENTS

RESULT 1  
AAAF89840  
ID AAAF89840 standard; DNA; 1128 BP.  
AC  
XX  
AC  
AAAF89840;  
XX  
DT 23-JUL-2001 (first entry)  
XX  
DE DNA encoding major merozoite surface protein-1 fragment of 42kDa.  
XX  
KW Major merozoite surface protein-1; MSP1-42; melittin signal peptide;  
XX malaria vaccine; ss.  
XX Plasmodium falciparum.  
XX  
FH Key Location/Qualifiers  
CDS 1..1128  
FT /\*tag= a

FT /product= "major merozoite surface protein-1 fragment  
of 42kDa"

WO200134188-A1.

17-MAY-2001.

09-NOV-2000; 2000WO-0531064.

12-NOV-1999; 99US-0165178.

01-DEC-1999; 99US-0168327.

22-AUG-2000; 2000US-0226861.

(UYHA-) UNIV HAWAII.

(UYCH-) UNIV CHINESE HONG KONG.

(QUEE-) QUEEN EMMA FOUND.

Hui GSN, Lap-Yin P, Ho WKK;

WPI; 2001-335879/35.

P-PSDB; AAB83926.

Producing malaria vaccine, useful for treatment or prevention of all forms of malaria in humans, by expressing immunogenic merozoite protein fragment in a baculovirus system

Example 3; Page 88-89; 95pp; English.

The present sequence encodes a major merozoite surface protein-1 C-terminal fragment of 42kDa (MSPL-42). This fragment is linked to a melittin signal peptide, and then expressed in a in a silk worm/baculovirus system. The protein is used to prepare a malaria vaccine, which is used to treat or prevent malaria, caused by any of the four species of Plasmodium that infect humans.

SQ Sequence 1128 BP; 487 A; 149 C; 163 G; 329 T; 0 other;

Alignment Scores:

Pred. No.: 1,23e-155 Length: 1128  
Score: 1970.00 Matches: 375  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 86.59% Indels: 0  
DB: 22 Gaps: 0

US-10-057-531A-2 (1-431) x AAF89840 (1-1128)

QY 57 MetAlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspVal 76  
DB 1 ATGCCATATCTGTCACATGGATATATCTCTCAGGATTTGAAATGATATGATGTT 60  
QY 77 IleTyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsn 96  
DB 61 ATATATTTAAACCTTTAGCTGGAGTATATAGAACGCTTAAACCAACAAATGAAAAAAC 120  
QY 97 IlePheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLys 116  
DB 121 ATTTTACATTTAATTTAAATTTGACGATATCTTAAATTCACGCTTAAAGAACGAAAA 180  
QY 117 TyrPheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGlu 136  
DB 181 TATTTCTTAGATGATTAGAACTGATTTAATGCAATTTAAACATATATCTCAATGAA 240  
QY 137 TyrIleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLys 156  
DB 241 TACATTATTGAAGATTCATTTAATTTAATTTGACACCAACAAAAACACACITTTAAAA 300  
QY 157 SerTyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIle 176  
DB 301 AGTTACAAATATATAAAGAAATCAGTAAATGATATTAATTTGACACAGGAGGTATA 360  
QY 177 SerTyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysVal 196

Db 361 AGTTATTTATGAAAGGTTTGTAGCGAAATATAAGGATGATTTAGAAATCAATTAATAAGTT 420  
QY 197 IleLysGluGluLysGluLysPheProSerSerProThrThrProSerProAla 216  
Db 421 ATCAAGAGAGAGAGAGAGAGTTCCTCATCATCACCACCAACACACCTCCGTCCACAGCA 480  
QY 217 LysThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThr 236  
Db 481 AAAACAGACGACCAACAAAGAGAAAGTAAGTTCCTCCATTTTAAACAAACATTTGAGACC 540  
QY 237 LeuTyrAsnAsnLeuValAsnLysIleAspAspTyrIleLeuLeuAsnLeuLysAlaLysIle 256  
Db 541 TTATACAAATACTTAGTTAATAAATTCAGCATTTACTTAATTAACCTTAAAGGCAAGATT 600  
QY 257 AsnAspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeu 276  
Db 601 ACGATTGTAAATGTTGAAAAGATGAAGCACATGTTAAAATACTAACTTAGTGATTTA 660  
QY 277 LysAlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLys 296  
Db 661 AAAGCAATTGATGACAAAATAGATCTTTTAAAAACCTTACGACTTCGAAGCAATTTAA 720  
QY 297 LysLeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuSerThrGlyLeu 316  
Db 721 AAATTGATAATGATGATACGAAAAAGATATGCTTGGCAAAATTTACTTAGTACAGGATTA 780  
QY 317 ValGlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMet 336  
Db 781 GTTCAAAATTTTCTTATACATAATCAATAATCAAAATTAATTCAGGAAATTTCCAAGATATG 840  
QY 337 LeuAsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPhe 356  
Db 841 TTAACATTTTCAACAACCAATCGTAAAAACAATGTCAGAAAAATTTCTGGATGTTTC 900  
QY 357 ArgHisLeuAspGluArgGluGluCysLysCysLeuAsnTyrLysGlnGluGlyAsp 376  
Db 901 AGACATTTAGTAGAAGAGAGAAATGTAATGTTTAAATTTACAAACAGAGGTTGAT 960  
QY 377 LysCysValGluAsnProAsnProThrCysAsnGluAsnGlyGlyCysAspAlaAsp 396  
Db 961 AAATGTGTTCAAAATCCAAATCCTACTTGTAAACGAAATAATGTTGATGTCAGAT 1020  
QY 397 AlaThrCysThrGluGluAspSerGlySerSerArgLysLysIleThrCysGluCysThr 416  
Db 1021 GCCACATGTACCGAAGAGATTCAGGTAGCAGCAGAGAGAAATTCACATGTGAATGTACT 1080  
QY 417 LysProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431  
Db 1081 AAACCTGATCTTATCCACTTTTCGATGGTATTTTCTGCAGTTCC 1125  
RESULT 2  
AAQ80911  
ID AAQ80911 standard; cDNA; 5181 BP.  
XX  
AC AAQ80911;  
XX  
DT 24-AUG-1995 (first entry)  
XX  
DE Plasmodium falciparum MSA-1 gene cDNA.  
XX  
KW Plasmodium falciparum MSA-1 gene; recombinant poxvirus;  
KW multicomponent multistage malarial vaccines; immunogens;  
KW malaria diagnosis; ss.  
XX  
OS Plasmodium falciparum (p486195).  
XX  
PN WO9428930-A.  
XX  
PD 22-DEC-1994.  
XX  
PF 10-JUN-1994; 94WO-US06652.  
XX  
PR 11-JUN-1993; 93US-0075783.







```

FT      /*tag= b
FT      /note= "Plasmodium falciparum MSA-1 peptide"
FT      1894..1950
FT      /*tag= c
FT      /note= "anchor sequence added to improve
FT      immunogenecity"
XX
PN      W09726911-Al.
XX
XX      31-JUL-1997.
XX
XX      29-JAN-1997; 97WO-US01395.
XX
XX      29-JAN-1996; 96US-0593006.
XX
XX      (GEU ) UNIV GEORGETOWN.
XX
XX      Davidson EA, Yang S;
XX
XX      WPI; 1997-393372/36.
XX
XX      Malaria vaccine - comprises expression vector expressing fragment of
XX      merozoite surface antigen
XX
XX      Claim 33; Fig 2; 75pp; English.
XX
XX      The present sequence encodes a chimeric protein that contains a
XX      mammalian signal and anchor sequence, and a merozoite surface
XX      antigen-1 (MSA-1) carboxy terminal peptide. The C-terminal fragment of
XX      MSA-1 provides a more specific response than the complete MSA-1,
XX      and attachment of anchor and signal sequences improve the immunogenicity
XX      of the protein better than the use of an adjuvant. The chimeric protein,
XX      and expression vectors (analogues that express MSA-1 without either
XX      signal or anchor peptides), particularly in the form of recombinant
XX      vaccinia virus, are used in vaccines to prevent or treat malaria caused
XX      by Plasmodium falciparum. The vaccinia vector expresses the antigen
XX      fragment for many days, or even years, generating a long-lasting immune
XX      response (humoral and/or cell-mediated) against the merozoite form of
XX      the parasite, in humans or other animals.
XX
XX      Sequence 1950 BP; 830 A; 236 C; 277 G; 607 T; 0 other;
SQ
Alignment Scores:
Pred. No.:      7.29e-89      Length:      1950
Score:          1175.00      Matches:      226
Percent Similarity: 71.43%      Conservative: 59
Best Local Similarity: 56.64%      Mismatches: 90
Query Match:      52.65%      Indels:      24
DB:              16          Gaps:         4
US-10-057-531A-2 (1-431) x AAT97956 (1-1950)
QY      34 ProAspLeuGlyThrAspAspAspLysAlaMetAlaAspIleGlySerIleGluGly 53
DB      760 CCTATATTGGAGATCCGAGAGAT-----TATGATGATTAGGCAAGTAGTAACA 813
QY      54 ArgGlyThrMetAlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGlu 73
DB      814 GCAGAGACGAGTAACCTCTCCGCTA---ATTGATAACATACCTTCTATAAATTGAAAAATGAA 870
QY      74 TyrAspValIleTyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysGlnIle 93
DB      871 TATGAGGTTTATATTAAACCTTTAGCAGGTTTATAGAGGTTTAAAAAACAATTA 930
QY      94 GluLysAsnIlePheThrPheAsnLeuAsnLeuAspIleLeuAsnSerArgLeuLys 113
DB      931 GAAATAACGTTATGACATTTATGTTATGTTAGGATATTATTAATTCACGATTTAAT 990
QY      114 LysArgLysTyrPheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSer 133
DB      991 AAACGTGAAATTTCAAAAATGTTTAGAATCAGATTTAATTCATATAAAGATTTAACA 1050
QY      134 SerAsnGluTyrIleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThr 153

```

```

DB      1051 TCAAGTAATATGTTCTCAAGATCCATATAAATTTCTTAATAACAAAGAGATAAA 1110
QY      154 LeuLeuLysSerTyrIleLysGluSerValGluAsnAspIleLysPheAlaGln 173
DB      1111 TTTCTTAAGCAGTTATAATATTAAGGATTCATAGATACGATGATATAAATTTGCAAT 1170
QY      174 GluGlyIleSerTyrGluLysValLeuAlaLysValLeuLysAspAspLeuGluSerIle 193
DB      1171 GATGTTCTTGGATATTATAAATATATTCGAAAAATATAAATCAGATTTAGATTCAT 1230
QY      194 LysLysValIleLysGluGluLysPheProSerSerProProThrProPro 213
DB      1231 AAAAAATATATC----- 1242
QY      214 SerProAlaLysThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsn 233
DB      1243 -----AACGCAAAACAAAGGTGAAATGAGAAATACCTCCCTTTTAAACAAT 1290
QY      234 IleGluThrLeuTyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLys 253
DB      1291 ATTGACACCTTATATAAACAGCTTAATGATAAATTCATTATTGTAATTCATTTAGAA 1350
QY      254 AlaLysIleAsnAspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeu 273
DB      1351 GCAAAAGTTCTAAATTTATACATATGAGAAATCAACGCTAGAACTTAAATAAAGAACTT 1410
QY      274 SerAspLeuLysAlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGlu 293
DB      1411 AATTACTTAAAAACAATTCAGACAAATTTGGCAGATTTTAAAAAATAAACATTTTCGTT 1470
QY      294 AlaIleLysLysLeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuSer 313
DB      1471 GGAATTTGCTGATTTATCAACAGATTATACCATTAATACTTATGACAAAGTTCCTTAGT 1530
QY      314 ThrGlyLeuVal---GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLys 332
DB      1531 ACAGGTATGTTGTTTGAATACTTGTCTAAACCGTTTATCTAATTTACTTGATGGAAC 1590
QY      333 PheGlnAspMetLeuAsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsn 352
DB      1591 TTGCAAGGTATGTTAAACATTTTCAACACCACTGCTGAAAAAACAATGTCACAAAAT 1650
QY      353 SerGlyCysPheArgHisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLys 372
DB      1651 TCTGGATGTTTCAGACATTTAGATGAAGAGAGATGTAATGTTTATTAATTTACAA 1710
QY      373 GlnGluGlyAspLysCysValGluAsnProAsnProThrCysAsnGluAsnGlyGly 392
DB      1711 CAAGAAGGTGATAAATGTTGTAATAATCCAAATCCTACTTGTAAACGAAAAATAATGTTGA 1770
QY      393 CysAspAlaAspAlaThrCysThrGluAspSerGlySerSerArgLysLysIleThr 412
DB      1771 TGTGATGCAAGATGCCAAATGTACCGAAGAGATTTCAGGTAGCAACGAAAGAAATTCACA 1830
QY      413 CysGluCysThrLysProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431
DB      1831 TGTGAATGTACTAAACCTGATTTCTTATCCACTTTTTCGATGTTTCTTCGAGTTCC 1887
RESULT 5
AAV21451
ID      AAV21451 standard; DNA; 4940 BP.
XX
XX      AAV21451;
XX
XX      23-SEP-1998 (first entry)
XX
XX      P. falciparum modified gp190 DNA.
XX
XX      gp190; malaria; MSP-1; merozoite surface protein; stability; vaccine;
XX      monoclonal antibody; passive immunisation; parasite; ss.
XX
XX      Plasmodium falciparum.

```



FH Key Location/Qualifiers  
 FT misc\_feature 1..165  
 FT /\*tag= a  
 FT /note= "signal sequence added to improve  
 FT immunogenicity"  
 FT  
 FT misc\_feature 166..1896  
 FT /\*tag= b  
 FT /note= "Plasmodium falciparum MSA-1 peptide"  
 FT  
 PN WO9726911-Al.  
 XX  
 XX  
 PD 31-JUL-1997.  
 XX  
 XX 29-JAN-1997; 97WO-US01395.  
 XX  
 XX 29-JAN-1996; 96US-0593006.  
 XX  
 XX (GEOU ) UNIV GEORGETOWN.  
 PA  
 PI Davidson EA, Yang S;  
 XX  
 XX WPI; 1997-393372/36.  
 XX  
 XX Malaria vaccine - comprises expression vector expressing fragment of  
 PT merozoite surface antigen  
 XX  
 XX Claim 33; Fig 3; 75pp; English.  
 XX  
 CC The present sequence encodes a chimeric protein that contains a mammalian  
 CC signal sequence, and a merozoite surface antigen-1 (MSA-1) carboxy  
 CC terminal peptide. The C-terminal fragment of MSA-1 provides a more  
 CC specific response than the complete MSA-1, and attachment of the  
 CC signal sequence improve the immunogenicity of the protein better than  
 CC the use of an adjuvant. The chimeric protein, and expression vectors  
 CC (analogues that express MSA-1 without either signal or anchor peptides),  
 CC particularly in the form of recombinant vaccinia virus, are used in  
 CC vaccines to prevent or treat malaria caused by Plasmodium falciparum. The  
 CC vaccinia vector expresses the antigen fragment for many days, or even  
 CC years, generating a long-lasting immune response (humoral and/or  
 CC cell-mediated) against the merozoite form of the parasite, in humans or  
 CC other animals.  
 XX  
 SQ Sequence 1896 BP; 813 A; 230 C; 272 G; 581 T; 0 other;

Alignment Scores:  
 Pred. No.: 1,25e-88 Length: 1896  
 Score: 1172.00 Matches: 225  
 Percent Similarity: 71.43% Conservative: 60  
 Best Local Similarity: 56.39% Mismatches: 90  
 Query Match: 51.52% Indels: 24  
 DB: 18 Gaps: 4

US-10-057-531A-2 (1-431) x AAT97957 (1-1896)

QY 34 ProAspLeuGlyThrAspAspAspAspPlyAlaMetAlaAspLeuGlySerLeuGly 53  
 DB 760 CCTATATTGGAGAAATCCGAAGAAGAT-----TATGATGATTTAGCAAGTAGTAACA 813  
 QY 54 ArgGlyThrMetAlaLeuSerValThrMetAspAsnLeuSerGlyPheGluAsnGlu 73  
 DB 814 GGAGAACGACTACTCTCCGTA---ATTGATAACATACACTTTCTAAATTTGAAATGAA 870  
 QY 74 TyrAspValIleTyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIle 93  
 DB 871 TATGAGGTTTATATTATAAACCTTTAGCAGGCTTTATAGAGTTTAAAAACACATTA 930  
 QY 94 GluLysAsnIlePheThrPheAsnLeuAsnLeuAsnAspLeuLeuAsnSerArgLeuLys 113  
 DB 931 GAAATAACGCTATGACATTTAATGTTAATGTTAAGGATATTTAAATTCACGATTTAAT 990  
 QY 114 LysArgLysTyrPheLeuAspValLeuSerAspLeuMetGlnPheLysHisIleSer 133  
 DB 991 AAACGTGAAATTTTCAAAATGTTTGAATTCAGATTTAATTCATATAAAGATTAAACA 1050

QY 134 SerAsnGluTyrIleLeuGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThr 153  
 DB 1051 TCAAGTAATTATGTTCAAGATCCATATAAATTTCTTAATAAGAAAAGAGATAAA 1110  
 QY 154 LeuLeuLysSerTyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGln 173  
 DB 1111 TTCTTAAGCACCTTATAATTATTAAGGATTCATAGATACGGATATAAATTTTGCAAT 1170  
 QY 174 GluGlyIleSerTyrTyrGluLysValLeuAlaLysTyrLysAspLeuGluSerIle 193  
 DB 1171 GATGTTCTTCGATATTATAAATATTATCCGAAATAATATAAATCAGATTTAGATTCAAT 1230  
 QY 194 LysLysValIleLysGluGluLysGluLysPheProSerProThrThrProPro 213  
 DB 1231 AAAAATATATC----- 1242  
 QY 214 SerProAlaLysThrAspGluGlnLysGluSerLysPheLeuProPheLeuThrAsn 233  
 DB 1243 -----AACGACAAACAGGTGAAATGAGAAATACCTTCCTTTTAAACAAT 1290  
 QY 234 IleGluThrLeuTyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLys 253  
 DB 1291 ATTGAGACCTTATATAAACAGTTAATGATAAATTCATTATTGTAATTCATTAGAA 1350  
 QY 254 AlaLysIleAsnAspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeu 273  
 DB 1351 GCAAAAGTTCTAATTTATACATATGAGAAATCAACCTGAGAGTTAAATAAAGAACTT 1410  
 QY 274 SerAspLeuLysAlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGlu 293  
 DB 1411 AATTACTTAAAAACAATTCAGACAAATTCGCAGATTTAAAAAAAATAACAAATTCGTT 1470  
 QY 294 AlaIleLysLysLeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSer 313  
 DB 1471 GGAATTCCTGATTATCAACAGATTTAACCATAATACCTATTATGACAAAGTTCCCTTAGT 1530  
 QY 314 ThrGlyLeuVal---GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLys 332  
 DB 1531 ACAGGTATGTTTGTGAAATCTTGCTAAAAACCGTTTTATCTAATTTACTTGATGGAAC 1590  
 QY 333 PheGlnAspMetLeuAsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsn 352  
 DB 1591 TTGCAAGGTATGTTTAAACATTTTCAACACCAATTCGCTAAAAAACAATGTCACAAAT 1650  
 QY 353 SerGlyCysPheArgHisLeuAspGluArgGluCysLysCysLysLeuLeuAsnTyrLys 372  
 DB 1651 TCTGGATGTTTCAGACATTTAGATGCAAGAGAGAAATGTAATGTTTATTAAATACAAA 1710  
 QY 373 GlnGluGlyAspLysCysValGluAsnProAsnProThrCysAsnGluAsnAsnGlyGly 392  
 DB 1711 CAAGAAGGTGATAAATGTTGAAATCCAAATCCTACTTCTTAACGAAATAATAGTGA 1770  
 QY 393 CysAspAlaAspAlaThrCysThrGluGluAspSerGlySerSerArgLysLysIleThr 412  
 DB 1771 TGTGATCCAGATGCAAAATGTTACCGAAAGAGATTTCAGGTAGCAACGGAAGAAATCACA 1830  
 QY 413 CysGluCysThrLysProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431  
 DB 1831 TGTGAATGTAATAAACCTGATTCTTATCCACTTTTCGATGTTATTTTCTGCAGTTCC 1887

RESULT 7  
 AAX56008  
 ID AAX56008 standard; cDNA; 1065 BP.  
 XX  
 AC AAX56008;  
 XX 31-AUG-1999 (first entry)  
 XX Merozoite surface protein MSP-1-42 modified cDNA.  
 DE  
 KW MSP-1; merozoite surface protein; malaria; vaccine;  
 KW protein engineering; codon usage; transgenic animal; ss.

XX Plasmodium falciparum.  
OS Synthetic.  
XX WO9920774-A2.  
XX PD 29-APR-1999.  
XX 20-OCT-1998; 98WO-US22226.  
XX 15-MAY-1998; 98US-0085649.  
PR 20-OCT-1997; 97US-0062592.  
XX (GENZ ) GENZYME TRANSGENICS CORP.  
XX Chen LH, Meade H;  
PI WPI; 1999-288313/24.  
DR P-PSDB; AAY09372.  
XX Modified malarial protein for use in anti-malarial vaccines  
XX Claim 1; Fig 1; 35pp; English.  
XX This novel, modified nucleic acid encodes the 42 kDa C-terminal  
CC portion (see AAY09372) of malaria merozoite surface protein MSP-1  
CC (MSP-1-42), an important target for the development of a vaccine  
CC against Plasmodium falciparum. The nucleic acid sequence has been  
CC modified compared to the native sequence of MSP-1-42 (see AAX56009)  
CC such that 306 nucleotide positions have been replaced to lower the  
CC AT content (from 76 to 49.7%) and to eliminate 10 mRNA instability  
CC motifs while maintaining the same protein amino acid sequence.  
CC These alterations allow MSP-1-42 to be expressed in mammalian cell  
CC culture and in transgenic mice. Native MSP-1-12 is known to be  
CC difficult to express in cell culture systems, mammalian cell  
CC culture systems or in transgenic animals. The invention allows  
CC expression of MSP-1-42 protein in the milk of transgenic animals,  
CC and also provides a DNA vaccine comprising a vector containing the  
CC altered MSP-1-42 sequence.  
XX SQ Sequence 1065 BP; 324 A; 254 C; 280 G; 207 T; 0 other;

Alignment Scores:  
Pred. No.: 4,71e-88 Length: 1065  
Score: 1161.50 Matches: 218  
Percent Similarity: 73.71% Conservative: 54  
Best Local Similarity: 59.08% Mismatches: 76  
Query Match: 51.05% Indels: 21  
DB: 20 Gaps: 2

US-10-057-531A-2 (1-431) x AAX56008 (1-1065)

Qy 63 MetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIleTyrLeuLysProLeu 82  
Db 19 ATCGATAACATCTCTCCAAAGATCGAAGACGAGTACGAGGTGCTGTACCTGAAGCGCGTG 78  
Qy 83 AlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIlePheThrPheAsnLeu 102  
Db 79 GCAGGGGTCTACCGGAGCGCTGAAGAAGACGCTGGAGAACACGCTGATGACCTTCAACGTG 138  
Qy 103 AsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyrPheLeuAspValLeu 122  
Db 139 AACGTGAAGGATATCTCTCAACAGCGCGTTCAACAGCGGAGAACTTCAAGACGTGCTG 198  
Qy 123 GluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyrIleIleGluAspSer 142  
Db 199 GAGACGATCTGATCCCTACAAAGATCTGACACGACGACCACTAGTGTGTCAAGGATCC 258  
Qy 143 PheLysLeuLeuAsnSerGlnLysAsnThrLeuLysSerTyrLysTyrIleLys 162  
Db 259 TACAGTTCTTCAACAGGAGAGATAAGTTCTCTGAGCAGGTACAACTACATCAAG 318  
Qy 163 GluSerValGluAsnAspIleLysPheAlaGlnGluLysSerTyrTyrGluLysVal 182

Db 319 GATAGCATGTATACCGGATATCAACTTCGCCAACAGATGTCTCTGGGATACTACAAAGATCTCTG 378  
Qy 183 LeuAlaLysTyrLysAspLeuGluSerIleLysValIleLysValIleLysLysGluLysGlu 202  
Db 379 TCCGAGAAGTACAAGAGCGATCTGGATTCAATCAAGAGTACATC----- 423  
Qy 203 LysPheProSerSerProThrThrProProSerProAlaLysThrAspGluGlnLys 222  
Db 424 -----AACGATAAGCAGGA 438  
Qy 223 LysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrAsnAsnLeuVal 242  
Db 439 GAGAACGAGAGTACCTGCTCTTCTGATCAACATCTGAGACCTGTACAAAGACCTGTCAAC 498  
Qy 243 AsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsnValGlu 262  
Db 499 GATAAGATTGATCTGCTGATCCACCTGGAGGCCAAGGTCTCTGAACCTACATATGAG 558  
Qy 263 LysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLysAlaLysAspLys 282  
Db 559 AAGAGCAACGTGGAGTCAAGATCAAGAGCTGAATTACCTGAAGACCATCCAGGATAAG 618  
Qy 283 IleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLysLeuIleAsnAsp 302  
Db 619 CTGGCCGATTTCAGAGAACAACAACACTTCGTGGGATCGCCGATCTGAGCACCGATTAC 678  
Qy 303 ThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal---GlnAsnPhePro 321  
Db 679 AACCAACAACCTCTCTGACCAAGTTCCTGAGCACCGGTATGCTTCCAAAACCTGGCC 738  
Qy 322 AsnThrIleLeuSerLysLeuIleGluLysPheGlnAspMetLeuAsnIleSerGln 341  
Db 739 AAGACCGTCTGAGCAACCTGCTGGATGGAACTGCGAGGGATCTGAACATCAGCCAG 798  
Qy 342 HisGlnCysValLysLysCysProGluAsnSerGlyCysPheArgHisLeuAspGlu 361  
Db 799 CACCAAGTGTGAAGAGCAGTGTCCCAAGACACGGGTGTTCAGACACCTGGATGAG 858  
Qy 362 ArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluLysAspLysCysValGluAsn 381  
Db 859 AGAGAGGAGTGTAAAGTGTCTGTAACCTACAAGCAGGAAGGTGATAAGTGTGCAAAAC 918  
Qy 382 ProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAlaThrCysThrGlu 401  
Db 919 CCAATCTCTACTTGTACGAGAACAAATGCTGGATGTGATGCCGATCCCAAGGTACCGAG 978  
Qy 402 GluAspSerGlySerSerArgLysLysIleThrCysGluCysThrLysProAspSerTyr 421  
Db 979 GAGGATTCAGGAGCAGCGGAGAGAGATCACCTGTGAGTGTACCAAGCCTGATCTTAT 1038

Qy 422 ProLeuPheAspGlyIlePheCysSer 430  
Db 1039 CCATGTTTCGATGGTATCTTCTGTAGT 1065

RESULT 8  
AAX25586  
ID AAX25586 standard; cDNA; 1065 BP.  
XX  
AC AAX25586;  
XX  
DT 02-AUG-1999 (first entry)  
XX  
DE Merozoite surface protein MSP-1-42 modified cDNA.  
XX  
KW MSP-1: merozoite surface protein; malaria; vaccine;  
KW protein engineering; protein expression; codon usage;  
KW transgenic animal; ss.  
OS Plasmodium falciparum.  
OS Synthetic.  
XX  
PN WO9920766-A2.

XX PD 29-APR-1999.  
XX PF 20-OCT-1998; 98WO-US22225.  
XX PR 15-MAY-1998; 98US-0085649.  
XX PR 20-OCT-1997; 97US-0062592.  
XX PA (GENZ ) GENZYME TRANSGENICS CORP.  
XX PI Chen LH, Meade H;  
XX DR WPI; 1999-302742/25.  
XX DR P-PSDB; AAY05832.  
XX PT New modified recombinant nucleic acid sequences useful for producing  
XX PT malarial DNA vaccine  
XX PS Claim 12; Fig 1; 43pp; English.  
XX CC This novel, modified nucleic acid encodes a 42 kDa C-terminal  
CC part (see AAY05832) of malaria merozoite surface protein MSP-1  
CC (MSP-1-42), an important target for the development of a vaccine  
CC against Plasmodium falciparum. The nucleic acid sequence has been  
CC modified compared to the native sequence of MSP-1-42 (see AAX25587)  
CC such that 306 nucleotide positions have been replaced to lower the  
CC AT content (from 76 to 49.7%) and to eliminate 10 mRNA instability  
CC motifs while maintaining the same protein amino acid sequence.  
CC These alterations allow MSP-1-42 to be expressed in mammalian cell  
CC culture and in transgenic mice. The invention provides modified  
CC recombinant nucleic acid sequences and methods for increasing the  
CC mRNA levels and protein expression of proteins that are difficult  
CC to express in cell culture systems, mammalian cell culture systems  
CC or in transgenic animals. The preferred difficult protein  
CC candidates for expression are those derived from lower organisms  
CC such as parasites, bacteria and viruses that have DNA coding  
CC sequences of high AT content or which have mRNA instability motifs  
CC or rare codons relative to the recombinant expression system to be  
CC used. The invention allows expression of MSP-1 protein in the milk  
CC of transgenic animals, and also provides a DNA vaccine comprising a  
CC vector containing the altered MSP-1-42 sequence.  
XX SQ Sequence 1065 BP; 324 A; 254 C; 280 G; 207 T; 0 other;  
  
Alignment Scores:  
Pred. No.: 4,71e-88 Length: 1065  
Score: 1161.50 Matches: 218  
Percent Similarity: 73.71% Conservative: 54  
Best Local Similarity: 59.08% Mismatches: 76  
Query Match: 51.05% Indels: 21  
DB: 20 Gaps:  
  
US-10-057-531A-2 (1-431) x AAX25586 (1-1065)  
  
QY 63 MetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIleTyrLeuLeu 82  
Db 19 ATCGATAACATCTGTCACAGATCGAGAACGAGTACGAGGTGCTGACCTGAAGCCGCTG 78  
  
QY 83 AlaGlyValTyrArgSerLeuLeuLysLysGlnIleGluLysAsnIlePheThrPheAsnLeu 102  
Db 79 GCAGGGGTCTACCGGAGCCTGAGAGAGCAGCTGGAGAACACGCTGATGACCTTCACAGTG 138  
  
QY 103 AsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyrPheLeuAspValLeu 122  
Db 139 ACGTGAAGGATATCTGTAACAGCCCGGTTCACAAAGCGGAGAACTTCAAGAACGCTGCTG 198  
  
QY 123 GluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyrIleIleGluAspSer 142  
Db 199 GAGAGCGATCTGATCCCTCCAGAGGATCTGACACGAGCAACTACGTGTCAAGGATCCC 258  
  
QY 143 PheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSerTyrIleLys 162  
Db 259 TACAAGTCTCTGAACAAGAGAGAGAGATAAGTCTCTGAGCAGTTACAACTACATCAAG 318

QY 163 GluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSerTyrTyrGluLysVal 182  
Db 319 GATAGCATTCATCCGATATCACTTCGCCAACGATCTCTGGGATCTACAGATCTCTG 378  
  
QY 183 LeuAlaLysTyrLysAspAspLeuGluSerIleLysValIleLysGluLysGlu 202  
Db 379 TCCGAGAAGTACAAGAGCGATCTGGATTCAATCAAGAAGTACATC----- 423  
  
QY 203 LysPheProSerSerProThrThrProProSerProAlaLysThrAspGluLys 222  
Db 424 -----AACGATAAGCAGGGA 438  
  
QY 223 LysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrAsnAsnLeuVal 242  
Db 439 GAGAACGAGAGTACCTGCTGCTGATCCACCTGAGGCGCAAGTCTCTGAACATACATATGAG 498  
  
QY 243 AsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsnValGlu 262  
Db 499 GATAAGATTGATCTGTTGATCCACCTGAGGCGCAAGTCTCTGAACATACATATGAG 558  
  
QY 263 LysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLysAlaIleAspAspLys 282  
Db 559 AAGAGCAACCTGGAGTCAAGATCAAGAGCTGAATTACCTCTGAAGACCATCCAGGTAAG 618  
  
QY 283 IleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLysLeuIleAsnAspAsp 302  
Db 619 CTGCGCGATTTCAGAGAACACAACTCTCTGCGGATCGCGATCTGAGCACCAGTATAC 678  
  
QY 303 ThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal----GlnAsnPhePro 321  
Db 679 AACCAACAAACCTGCTGACCAAGTCTCTGAGCACCAGTATGCTCTTCGAAACCTGGCC 738  
  
QY 322 AsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeuAsnIleSerGln 341  
Db 739 AACACCTCTCTGAGCAACCTCTGATGGAGAACCTGCGAGGGGATCTGAACATCAGCCAG 798  
  
QY 342 HisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArgHisLeuAspGlu 361  
Db 799 CACCAGTGTGTGAAGAAGCAGTGTCCCCAGAACAGCGGGTGTTCACACACACCTGGATGAG 858  
  
QY 362 ArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLysCysValGluAsn 381  
Db 859 AGAGAGGAGTGTAAAGTGTCTGCTGAACCTACAAAGCAGGAAGGTGATAAGTGTGTGAAAC 918  
  
QY 382 ProAsnProThrCysAsnGluAsnAsnGlyCysAspAlaAspAlaThrCysThrGlu 401  
Db 919 CCCAATCTTACTTGTACAGAGAACATGTTGGATGTGATGCCGATGCCAAGTGTACCCAG 978  
  
QY 402 GluAspSerGlySerArgLysLysIleThrCysGluCysThrLysProAspSerTyr 421  
Db 979 GAGGATTTCAGGAGCAACGCGGAGGAAGATCACCTGTGAGTGTACCAAGCCTGATTCTTAT 1038  
  
QY 422 ProLeuPheAspGlyIlePheCysSer 430  
Db 1039 CCAGTGTTCGATGATCTCTCTAGT 1065  
  
RESULT 9  
AAC68976  
ID AAC68976 standard; DNA; 1077 BP.  
XX AAC68976;  
XX 27-FEB-2001 (first entry)  
XX Merozoite surface protein-142 coding sequence.  
XX Merozoite surface protein; protozoasome; vaccine; malaria; ss.  
XX Plasmodium falciparum.  
XX WO200063245-A2.  
XX



PR 20-OCT-1997; 97US-0062592.  
 XX (GENZ ) GENZYME TRANSGENICS CORP.  
 XX Chen LH, Meade H;  
 XX WPI: 1999-288313/24.  
 DR P-PSDB; AAY09373.  
 XX  
 PT Modified malarial protein for use in anti-malarial vaccines  
 XX  
 XX Example; Fig 2; 35pp; English.  
 XX  
 CC This nucleic acid encodes a 42 kDa C-terminal portion (see AAY09373)  
 CC of malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an  
 CC important target for the development of a vaccine against  
 CC Plasmodium falciparum. The 3' end of the sequence has been  
 CC modified to include a 6xHis tag. The nucleic acid sequence has  
 CC been modified (see AAX56008) according to a method of the invention  
 CC in order to improve expression in mammalian host cells and in  
 CC transgenic animals. In the modified coding sequence, 306 nucleotide  
 CC positions have been replaced to lower the AT content (from 76 to  
 CC 49.7%) and to eliminate 10 mRNA instability motifs. The encoded  
 CC amino acid sequence is unaltered. In another modified sequence  
 CC (see AAX56021), a signal peptide sequence has been added and two  
 CC N-glycosylation sites eliminated. The invention allows expression  
 CC of MSP-1-42 protein in the milk of transgenic animals, and also  
 CC provides a DNA vaccine comprising a vector containing the altered  
 CC MSP-1-42 sequence.  
 XX  
 SQ Sequence 1088 BP; 454 A; 139 C; 150 G; 345 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 4.83e-88 Length: 1088  
 Score: 1161.50 Matches: 218  
 Percent Similarity: 73.71% Conservative: 54  
 Best Local Similarity: 59.08% Mismatches: 76  
 Query Match: 51.05% Indels: 21  
 DB: 20 Gaps: 2  
 US-10-057-531A-2 (1-431) x AAX56009 (1-1088)  
 QY 63 MetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIleTyrLeuLysProLeu 82  
 DB 19 ATTGATAACATCTTCTCTAAATGAATGAATGACGAGGTTTATATATTAACACCTTTA 78  
 QY 83 AlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIlePheThrPheAsnLeu 102  
 DB 79 GCAGGTGTTATAGAGTTTAAATAAACAACTAGATAAATGACGTTATGACATTTAATGTT 138  
 QY 103 AsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyrPheLeuAspValLeu 122  
 DB 139 AATGTTAAGGATATTTAAATTTTAAATTAACAGTGAAATTTCAAAATGTTTAA 198  
 QY 123 GluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyrIleIleGluAspSer 142  
 DB 199 GAATCAGATTTAATTCATATAAGATTTAATCAATGAATATATGTTCAAGATCCA 258  
 QY 143 PheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSerTyrLysTyrIleLys 162  
 DB 259 TATAAATTTCTTAATAAGAAAGAAAGATATAATCTTGAAGCATTTATATATTAAG 318  
 QY 163 GluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSerTyrTyrGluLysVal 182  
 DB 319 GATTCAATAGATACGATATAATTTGCAATGATGTTCTTGATATATATAAATATTA 378  
 QY 183 LeuAlaLysTyrLysAspLeuGluSerIleLysValIleLysGluLysGlu 202  
 DB 379 TCCGAAAAATATAAATCAGATTAGATTCAATTAATAAATAATATATC----- 423  
 QY 203 LysPheProSerSerProThrThrProProSerProAlaLysThrAspGluGlnLys 222  
 DB 424 -----AACGACAAACAAGGT 438

QY 223 LysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrAsnAsnLeuVal 242  
 DB 439 GAAATGAGAAATACCTTCCTTTTAAACAATATTTGAGACCTTATATAAACAGTAAT 498  
 QY 243 AsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsnValGlu 262  
 DB 499 GATAAAATGATTTATTTGATTTCTTAAAGCAAAAGTTCTAAATATATACATATGAG 558  
 QY 263 LysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLysAlaIleAspAspLys 282  
 DB 559 AATCAACGTCAGAGTTAAATAAAGAACTTAACTTAAACAACTTAAAGCAAA 618  
 QY 283 IleAspLeuLysAsnProTyrAspPheGluAlaIleLysLysLeuIleAsnAspAsp 302  
 DB 619 TTGGCAGATTTTAAATAAATAAATAATTCGTTGGAATTCCTGATTTATCAACAGATTAT 678  
 QY 303 ThrLysLysAspMetLeuLysLeuLeuSerThrGlyLeuVal---GlnAsnPhePro 321  
 DB 679 AACCATTAATAACTATTGACAAAGTTCTTAGTACAGGTATGGTTTTTGAATAATCTTGTCT 738  
 QY 322 AsnThrIleIleSerLysLeuIleGlyLysPheGlnAspMetLeuAsnIleSerGln 341  
 DB 739 AAACCGTTTATCTTAATTTACTTTGATGAAGAACTTCGAAGGTATGTTAAACATTTCAAA 798  
 QY 342 HisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArgHisLeuAspGlu 361  
 DB 799 CACCAATGCTAAAAAACAATGTCACCAAAATTCCTGGATGTTTCAGACATTTAGATGAA 858  
 QY 362 ArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluLysAspLysCysValGluAsn 381  
 DB 859 AGAAGAAGATGAAATGTTTATTAATTAACAAAGAAAGGTGATAAATGTGTTGAAAT 918  
 QY 382 ProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAlaThrCysThrGlu 401  
 DB 919 CCAATCTCTACTGTAAACGAAATAATATGTTGATGTCAGATGCCAATGTACCGAA 978  
 QY 402 GluAspSerGlySerArgLysLysIleThrCysGluCysThrLysProAspSerTyr 421  
 DB 979 GAAGATTTCAGGTAGCAACGCAAAAGAAATCAGATGTGAATGTACTAAACCTGATTTCTAT 1038  
 QY 422 ProLeuPheAspGlyIlePheCysSer 430  
 DB 1039 CCACATTTTCGATGTTATTTCTGCAGT 1065  
 RESULT 11  
 AAX25587  
 ID AAX25587 standard; cDNA; 1088 BP.  
 XX  
 AC AAX25587;  
 XX  
 DT 02-AUG-1999 (first entry)  
 XX  
 DE Merozoite surface protein MSP-1-42 cDNA.  
 XX  
 KW MSP-1; merozoite surface protein; malaria; vaccine;  
 KW protein engineering; protein expression; codon usage;  
 KW transgenic animal; ss.  
 XX  
 OS Plasmodium falciparum.  
 FH Key Location/Qualifiers  
 CDS 1..1086  
 FT /\*tag= a  
 XX  
 PN WO9920766-A2.  
 XX  
 PD 29-APR-1999.  
 XX  
 PF 20-OCT-1998; 98WO-US22225.  
 XX  
 PR 15-MAY-1998; 98US-0085649.  
 PR 20-OCT-1997; 97US-0062592.

XX (GENZ ) GENZYME TRANSGENICS CORP.  
 XX Chen LH, Meade H;  
 XX WPI: 1999-302742/25.  
 DR P-PSDB; AAY05833.  
 DR  
 PT New modified recombinant nucleic acid sequences useful for producing  
 PT malarial DNA vaccine  
 PS Disclosure; Fig 2; 43pp; English.  
 XX  
 XX This nucleic acid encodes a 42 kDa C-terminal portion (see AAY05833)  
 CC of malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an  
 CC important target for the development of a vaccine against  
 CC Plasmodium falciparum. The 3' end of the sequence has been  
 CC modified to include a 6xHis tag. The nucleic acid sequence has  
 CC been modified (see AAX25586) according to a method of the invention  
 CC in order to improve expression in mammalian host cells and in  
 CC transgenic animals. In the modified coding sequence, 306 nucleotide  
 CC positions have been replaced to lower the AT content (from 76 to  
 CC 49.7%) and to eliminate 10 mRNA instability motifs. The encoded  
 CC amino acid sequence is unaltered. In another modified sequence  
 CC (see AAX25593), a signal peptide sequence has been added and two  
 CC N-glycosylation sites eliminated. The invention provides modified  
 CC recombinant nucleic acid sequences and methods for increasing the  
 CC mRNA levels and protein expression of proteins that are difficult  
 CC to express in cell culture systems, mammalian cell culture systems  
 CC or in transgenic animals. The preferred difficult protein  
 CC candidates for expression are those derived from lower organisms  
 CC such as parasites, bacteria and viruses that have DNA coding  
 CC sequences of high AT content or which have mRNA instability motifs  
 CC or rare codons relative to the recombinant expression system to be  
 CC used. The invention allows expression of MSP-1 protein in the milk  
 CC of transgenic animals, and also provides a DNA vaccine comprising a  
 CC vector containing the altered MSP-1-42 sequence.  
 XX  
 SO Sequence 1088 BP; 454 A; 139 C; 150 G; 345 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 4.83e-88 Length: 1088  
 Score: 1161.50 Matches: 218  
 Percent Similarity: 73.71% Conservative: 54  
 Best Local Similarity: 59.08% Mismatches: 76  
 Query Match: 51.05% Indels: 21  
 DB: 20 Gaps: 2  
 US-10-057-531A-2 (1-431) x AAX25587 (1-1088)  
 QY 63 MetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIleTyrLeuLysProLeu 82  
 Db ATTGATAACATACATCTTCTAAATTTGAAATTAATGAAATACGAGGTTTATATATTTAAACACCTTTA 78  
 QY 83 AlaGlyValTyrArgSerLeuLysGlnIleGluLysAsnIlePheThrPheAsnLeu 102  
 Db GCAGGTGTTTATACAGCTTTAAAAAACAATAGAAATACGTTATGACATTTAATGTT 138  
 QY 103 AsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyrPheLeuAspValLeu 122  
 Db AATGTTAAGGATATTTTAAATTCAGGATTTAATAAAGCTGAAATTTCAAAATGTTTAA 198  
 QY 123 GluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyrIleIleGluAspSer 142  
 Db GAATCAGATTTTAAATTCATTAAGATTTTAAACATCAAGTAAATATGTTGTCAAAGATCCA 258  
 QY 143 PheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSerTyrLysTyrIleLys 162  
 Db TATAAATTTCTTATAAAGAAAAAGAGATAAATCTTACGACGTTATATATATATTAAG 318  
 QY 163 GluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSerTyrTyrGluLysVal 182  
 Db GATTCAATAGATACGGATATATAAATTTTGCAAAATGATGCTTCTGGATATATATAAATATTA 378

QY 183 LeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValIleLysGluGluLysGlu 202  
 Db TCCGAAAAATATAATCAGATTAGATTCAATTAATAAATATATC----- 423  
 QY 203 LysPheProSerSerProProThrProProSerProAlaLysThrAspGluGlnLys 222  
 Db -----AACGACAAACAAGGT 438  
 QY 223 LysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrAsnAsnLeuVal 242  
 Db GAAATGAGAAATACCTTCCTTTTAAACAAATATGAGACCTTATATAAACAACGTTAAT 498  
 QY 243 AsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsnValGlu 262  
 Db GATAAATGATTATTTGTAATTCATTAGAACGAAAGTCTTAATATATCATATGAG 558  
 QY 263 LysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLysAlaIleAspLys 282  
 Db AATCAACACGTAGAGTTAAATTAAGAACTTAATTAATAAACAATTTCAAGACAAA 618  
 QY 283 IleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLysLeuIleAsnAsp 302  
 Db TTGCACATTTTAAAAAATAACATTTCTTGAATTCGCTGATTTATCAACAGATTAT 678  
 QY 303 ThrLysLysAspMetLeuGlyLysLeuSerThrGlyLeuVal---GlnAsnPhePro 321  
 Db AACCATATAACTATTGACAAAGTCTCTTAGTACAGGTATGGTTTTGAAAAATCTTGCT 738  
 QY 322 AsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeuAsnIleSerGln 341  
 Db AAAACCCGTTTATCTAATTTACTTTGAGAACTTGCAAGGTATGTTAAACATTTCAAA 798  
 QY 342 HisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArgHisLeuAspGlu 361  
 Db CACCAATCGGTAAACAAACAAATGTCACAAATCTCGATGTTTCAGACATTTAGATGAA 858  
 QY 362 ArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLysCysValGluAsn 381  
 Db AGAGAAGAATGTAATGTTTATTAAATACAAACAAGGTGATAATGTGTTGAAAAAT 918  
 QY 382 ProAsnProThrCysAsnGluAsnAsnGlyCysAspAlaAspAlaThrCysThrGlu 401  
 Db CCAATCTCTACTCTACGAAATAATGTTGGATGTTGATGAGATGCCAAATGTACCGAA 978  
 QY 402 GluAspSerGlySerSerArgLysLysIleThrCysGluCysThrLysProAspSerTyr 421  
 Db GAAGATTTCAGTACGACGAAAGAAATCACATGTCATGTAATGTAACCTGATTCATTAT 1038  
 QY 422 ProIlePheAspGlyIlePheCysSer 430  
 Db CCACCTTTTCGATGTTATTTCTGCACT 1065  
 RESULT 12  
 ID AAV35363  
 XX AAV35363 standard; DNA; 4940 BP.  
 AC AAV35363;  
 XX  
 DT 23-SEP-1998 (first entry)  
 XX  
 DE P. falciparum gp190 DNA.  
 XX  
 KW gp190; malaria; MSP-1; merozoite surface protein; stability; vaccine;  
 KW monoclonal antibody; passive immunisation; parasite; ss.  
 OS Plasmodium falciparum.  
 XX  
 FH Key Location/Qualifiers  
 CDS 10..4929  
 FT /\*tag= a  
 FT /product= gp190  
 XX











XX 26-OCT-2000.  
 XX 20-APR-2000; 2000WO-CB01558.  
 XX 20-APR-1999; 99GB-0009072.  
 PR 13-MAY-1999; 99US-0311817.  
 PR 25-MAY-1999; 99CA-2271451.  
 XX (MEDI-) MEDICAL RES COUNCIL.  
 XX Holder A, Birdsall B, Feeney J, Morgan W, Syed S, Uthaipibull C;  
 XX WPI; 2001-015762/02.  
 DR P-PSDB; AAB37610.  
 XX Novel variants of the C-terminal fragment of Plasmodium merozoite  
 PT surface protein-1, useful as vaccines for treating or preventing  
 PT malaria -  
 XX Example 5; Fig 15; 126pp; English.  
 XX The present invention relates to non-natural variants of a C-terminal  
 CC fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The  
 CC non-natural variants have reduced affinity for at least 1 antibody  
 CC capable of blocking a second antibody that inhibits the proteolytic  
 CC cleavage of Plasmodium MSP-1.4.2, and has the same affinity for at least  
 CC one third antibody that inhibits the proteolytic cleavage of Plasmodium  
 CC MSP-1.4.2, compared to natural MSP-1.1.9. The non-natural variants of the  
 CC present invention are useful for immunising a mammal against malaria, and  
 CC can be used to treat malaria. The present sequence is the MSP-133  
 CC coding sequence.  
 XX SQ Sequence 786 BP; 245 A; 164 C; 159 G; 218 T; 0 other;

Alignment Scores:  
 Pred. No.: 1.21e-44 Length: 786  
 Score: 640.50 Matches: 129  
 Percent Similarity: 65.22% Conservative: 51  
 Best Local Similarity: 46.74% Mismatches: 75  
 Query Match: 28.15% Indels: 21  
 DB: 22 Gaps: 2

US-10-057-531A-2 (1-431) x AAC68978 (1-786)

QY 63 MetAspAsnIleuSerGlyPheGluAsnGluTyrAspValIleTyrLeuLysProLeu 82  
 DB 19 ATCGATAACATCTCTCTAAGATTGAGACGATACGAGCTCTGTACTTGAGCCTCTG 78  
 QY 83 AlaGlyValTyrArgSerLeuLysGlnIleGluLysAsnIlePheThrPheAsnLeu 102  
 DB 79 GCCGGTGTCTACAGATCCCTGAAGAGCACTGGAACCAACGTCATGACTTTCACGTT 138  
 QY 103 AsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyrPheLeuAspValLeu 122  
 DB 139 AAGCTCAAGAGCATTTTGAACCTCCAGATTCACAGAGAGAACTTCAAGAACGTTCTG 198  
 QY 123 GluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyrIleIleGluAspSer 142  
 DB 199 GAGCTGACTTGTATCCATCAAGGATTTGACTTCTTCTTAACGTGTTAAGGCCCA 258  
 QY 143 PheLysLeuLeuAsnSerGlnLysAsnThrLeuLysSerTyrLysTyrIleLys 162  
 DB 259 TACAAGTCTCTCAACAGGAGAGAGACAAAGTTCTTCTCTCTTACAACTACATTAAAG 318  
 QY 163 GluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSerTyrTyrGluLysVal 182  
 DB 319 GACTCCATTTGACTATATCAACTTCTGCTAACGAGCTCTGGGTACTACAGATCTTG 378  
 QY 183 LeuAlaLysTyrLysAspLeuGluSerIleLysValIleLysGluLysGlu 202  
 DB 379 TCTGAGCAAGTACAGTCTGACTTGGATTCCATCAAGAGTATATC----- 423

QY 203 LysPheProSerProThrThrProProSerProAlaLysThrAspGluGlnLys 222  
 DB 424 -----NATGATAGCAAGGA 438  
 QY 223 LysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrAsnAsnLeuVal 242  
 DB 439 GAGAAATGAAAGTACCTGCTCATTTGGAAGCAAGGCTTGGCAATACACTTACAGAA 498  
 QY 243 AsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsnValGlu 262  
 DB 499 GATAGATCGATCTGTGCTCATTTGGAAGCAAGGCTTGGCAATACACTTACAGAA 558  
 QY 263 LysAspGluAlaHisValLysIleThrLysLeuSerAspLysAlaIleAspAspLys 282  
 DB 559 AAGTCTAACCTCGAGGTCAAGATCAAGGAATTGAACCTACCTCAAGACCATTCAGGATAAG 618  
 QY 283 IleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLysLeuIleAspAsp 302  
 DB 619 CTGGCTGATTTCAAGAAAGCAACAACTTCGTTGGTATCGCTGATTGTCCACTGATTAC 678  
 QY 303 ThrLysLysAspMetLeuGlyLysLeuSerThrGlyLeuVal--GlnAsnPhePro 321  
 DB 679 AACCAACAACATCTGTGACTAAGTCTCTACCGGTATGGTTTCGAGAACTTGGCT 738  
 QY 322 AsnThrIleLysSerLysLeuIleGlyLysPheGlnAspMetLeu 337  
 DB 739 AAGACTGTCTCTCAACCTGTTGGATGCTAACTTCAGGGTATGTTG 786  
 RESULT 17  
 AAT80403  
 ID AAT80403 standard; DNA; 354 BP.  
 XX  
 AC AAT80403;  
 XX  
 DT 25-MAR-1998 (first entry)  
 XX  
 DE PfMSP1(p19)A coding sequence.  
 XX  
 KW Chimeric; Plasmodium vivax; merozoite surface protein; MSP1; p19;  
 KW Plasmodium falciparum; malaria; vaccine; immunity; epitope; ss.  
 XX  
 OS Plasmodium falciparum.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..354  
 FT /tag= a  
 FT /product= "PfMSP1(p19)A with GPI anchoring sequence"  
 FT /note= "sequence contains 2 in frame stop codons at  
 the 3'-end"  
 FT misc\_feature 1..285  
 FT /tag= b  
 FT /note= "derived from P. falciparum MSP1 p19 sequence"  
 FT misc\_feature 286..354  
 FT /tag= c  
 FT /note= "glycosylphosphatidylinositol anchoring sequence"  
 XX WO9730159-A2.  
 XX  
 XX  
 PD 21-AUG-1997.  
 XX  
 PF 14-FEB-1997; 97WO-FR00291.  
 XX  
 XX 14-FEB-1996; 96FR-0001821.  
 XX  
 PA (INSP ) INST PASTEUR.  
 PA (UUNY ) UNIV NEW YORK STATE.  
 XX  
 PI Barnwell JW, Longacre-Andre S, Mendis K, Nato F;  
 PI Roth C;  
 XX  
 DR WPI; 1997-425034/39.  
 DR P-PSDB; AAW22592.



```

XX AAT80404;
AC 25-MAR-1998 (first entry)
DT
XX PfMSPl(p19)S coding sequence.
DE
XX Chimeric; Plasmodium vivax; merozoite surface protein; MSP1; p19;
KW Plasmodium falciparum; malaria; vaccine; immunity; epitope; ss.
XX
OS Chimeric - Plasmodium vivax.
OS Chimeric - Plasmodium falciparum.
XX
FH Key Location/Qualifiers
FT CDS 1..387
FT /tag= a
FT /product= PfMSPl(p19)S
FT /note= "sequence contains 2 in frame stop codons at
FT the 3' end"
FT
FT sig_peptide 1..57
FT /tag= b
FT mat_peptide 58..381
FT /tag= c
FT misc_feature 1..96
FT /tag= d
FT /note= "sequence derived from P. vivax MSP1"
FT misc_feature 97..102
FT /tag= e
FT /note= "sequence derived from generated restriction
FT enzyme site"
FT
FT misc_feature 103..381
FT /tag= e
FT /note= "sequence derived from P. falciparum p19 coding
FT sequence"
FT
XX WO730159-A2.
PN
XX 21-AUG-1997.
PD
XX 14-FEB-1997; 97WO-FR00291.
XX
XX 14-FEB-1996; 96FR-0001821.
PR
XX (INSP ) INST PASTEUR.
PA
XX (UUNY ) UNIV NEW YORK STATE.
XX
XX Barnwell JW, Longacre-Andre S, Mendis K, Nato F;
PI Roth C;
XX
XX WPI; 1997-425034/39.
DR
XX P-PSDB; AAW22592.
DR
XX Recombinant protein containing Plasmodium merozoite surface
PT protein-1 p42 fragment - useful in antimalarial vaccines, also new
PT antibodies for diagnosis and protein purification
XX
XX Disclosure; Fig 1C; 85pp; French.
PS
XX This is the nucleotide sequence encoding a chimeric protein comprising
CC nucleotides encoding amino acids 1-32 of the Plasmodium vivax merozoite
CC surface protein 1 (MSP1) linked to the nucleotide sequence encoding the
CC 19 kD C-terminal fragment (p19) of MSP-1 from Plasmodium falciparum.
CC p19 is the C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
CC The recombinant protein can be used for the production of anti-malarial
CC vaccines, where the p19 fragment provides a high level of protective
CC immunity since it includes epitopes not presented in the p42 fragment.
XX
XX Sequence 387 BP; 116 A; 94 C; 96 G; 81 T; 0 Other;
SQ

```

## Alignment Scores:

```

Pred. No.: 1.59e-35 Length: 387
Score: 527.00 Matches: 91
Percent Similarity: 95.83% Conservative: 1

```

```

Best Local Similarity: 94.79% Mismatches: 4
Query Match: 23.16% Indels: 0
DB: 18 Gaps: 0
US-10-057-531A-2 (1-431) x AAT80404 (1-387)
QY 335 AspMetLeuAsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGly 354
|||
DB 94 GAGCAATTCACATCTCGCAGCAGCAATCGTGAAAAAACAATGTCCCGAGAACTCTGGC 153
QY 355 CysPheArgHisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGlu 374
|||||
DB 154 TGTTCAGACACTTGGCAGCAGAGAGAGAGAGTGTAAATGTCTGTAACACAGGAG 213
QY 375 GlyAspLysCysValGluAsnProAsnProThrCysAsnGluAsnGlyGlyCysAsp 394
|||||
DB 214 GCGCAACAGTGGTGAGAACCCCAACCGACCTGTAAACGAGAACCAACGGCGCTGTGAC 273
QY 395 AlaAspAlaThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGlu 414
|||||
DB 274 GCAGACGCCCAATGCACCGCAGGAGGACTCGGCAGCAGCAAGAAATCAGCTGTGAG 333
QY 415 CysThrLysProAspSerTyrProLeuPheAspGlyIlePheCysSer 430
|||||
DB 334 TGTACCAAAACCGGACTCGTACCGCTGTTCGACGGCATCTTCTGCAGC 381
RESULT 20
AAT94549
ID AAT94549 standard; DNA; 387 BP.
XX
AC AAT94549;
XX
DT 25-MAR-1998 (first entry)
XX
DE PfMSPl(p19)S coding sequence.
XX
KW Chimeric; Plasmodium vivax; merozoite surface protein; MSP1; p19;
KW Plasmodium falciparum; malaria; vaccine; immunity; epitope; ss.
XX
OS Chimeric - Plasmodium vivax.
OS Chimeric - Plasmodium falciparum.
XX
FH Key Location/Qualifiers
FT CDS 1..387
FT /tag= a
FT /product= PfMSPl(p19)S
FT /note= "sequence contains 2 in frame stop codons at
FT the 3' end"
FT
FT sig_peptide 1..57
FT /tag= b
FT mat_peptide 58..381
FT /tag= c
FT misc_feature 1..96
FT /tag= d
FT /note= "sequence derived from P. vivax MSP1"
FT misc_feature 97..102
FT /tag= e
FT /note= "sequence derived from generated restriction
FT enzyme site"
FT
FT misc_feature 103..381
FT /tag= e
FT /note= "sequence derived from P. falciparum p19 coding
FT sequence"
FT
XX WO730158-A2.
PN
XX 21-AUG-1997.
PD
XX 14-FEB-1997; 97WO-FR00290.
XX
XX 14-FEB-1996; 96FR-0001822.
XX
XX (INSP ) INST PASTEUR.
PA

```

PA (UUNY ) UNIV NEW YORK STATE.  
 XX Barnwell JW, Longacre-Andre S, Mendis K, Nato F;  
 PI Roth C;  
 XX WPI: 1997-425033/39.  
 DR P-PSDB: AAW36102.  
 XX  
 XX Recombinant protein containing the merozoite surface protein-1 p19  
 PT fragment - useful in anti-malarial vaccines, diagnosis and protein  
 PT purification  
 XX  
 XX Disclosure; Fig 1C; 85pp; French.  
 XX  
 CC This is the nucleotide sequence encoding a chimeric protein comprising  
 CC nucleotides encoding amino acids 1-32 of the Plasmodium vivax merozoite  
 CC surface protein 1 (MSP1) linked to the nucleotide sequence encoding the  
 CC 19 kD C-terminal fragment (p19) of MSP-1 from Plasmodium falciparum.  
 CC p19 is the C-terminal fragment of the 42 kD MSP1 from Plasmodium species.  
 CC The recombinant protein can be used for the production of anti-malarial  
 CC vaccines, where the p19 fragment provides a high level of protective  
 CC immunity since it includes epitopes not presented in the p42 fragment.  
 XX  
 SQ Sequence 387 BP; 116 A; 94 C; 96 G; 81 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 1.59e-35 Length: 387  
 Score: 527.00 Matches: 91  
 Percent Similarity: 95.83% Conservative: 1  
 Best Local Similarity: 94.79% Mismatches: 4  
 Query Match: 23.16% Indels: 0  
 DB: 18 Gaps: 0  
 US-10-057-531A-2 (1-431) x AAT94549 (1-387)  
 Qy 335 AspMetLeuAsnIleSerGlnHisGlnCysValLysGlnCysProGluAsnSerGly 354  
 Db 94 GACGAATTCACATCTCCGACACCAATGCGTGAATAAACAANTGTCGGAAGACTCTGGC 153  
 Qy 355 CysPheArgHisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGlu 374  
 Db 154 TGTTCAGACACTGGACGACAGAGAGAGTGAATCTCTGCTGAACACAGGAG 213  
 Qy 375 GlyAspLysCysValGluAsnProAsnProThrCysAsnGluAsnAsnGlyCysAsp 394  
 Db 214 GCGCACAAGTCGCTGGAGAACCCCAACCCGACCTGTATACGAGAACACGCGGCTGTGAC 273  
 Qy 395 AlaAspAlaThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGlu 414  
 Db 274 GCAGACGCCAATGCACCGAGGAGGACTCGGCGACGACGACGCAAGAAATCACGTGTGAG 333  
 Qy 415 CysThrLysProAspSerTyrProLeuPheAspGlyIlePheCysSer 430  
 Db 334 TGTACCAAAACCCGACTCGTACCCGCTGTTCACGCGCATCTTCTGCAGC 381  
 RESULT 21  
 AAC68977  
 ID AAC68977 standard; DNA; 333 BP.  
 XX  
 AC AAC68977;  
 XX  
 DT 27-FEB-2001 (first entry)  
 XX  
 DE Merozoite surface protein-119 coding sequence.  
 XX  
 KW Merozoite surface protein; protazoacide; vaccine; malaria; ss.  
 OS Plasmodium falciparum.  
 XX  
 PN WO200063245-A2.  
 XX  
 PD 26-OCT-2000.  
 XX

PF 20-APR-2000; 2000WO-GB01558.  
 XX  
 PR 20-APR-1999; 99GB-0009072.  
 PR 13-MAY-1999; 99US-0311817.  
 PR 25-MAY-1999; 99CA-2271451.  
 XX  
 PA (MEDI-) MEDICAL RES COUNCIL.  
 XX  
 XX Holder A, Birdsall B, Feeney J, Morgan W, Syed S, Uthaipibull C;  
 PI WPI: 2001-015762/02.  
 DR P-PSDB: AAB37609.  
 XX  
 PT Novel variants of the C-terminal fragment of Plasmodium merozoite  
 PT surface protein-1, useful as vaccines for treating or preventing  
 PT malaria.  
 XX  
 PS Example 5; Fig 15; 126pp; English.  
 XX  
 CC The present invention relates to non-natural variants of a C-terminal  
 CC fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The  
 CC non-natural variants have reduced affinity for at least 1 antibody  
 CC capable of blocking a second antibody that inhibits the proteolytic  
 CC cleavage of Plasmodium MSP-1\_4\_2, and has the same affinity for at least  
 CC one third antibody that inhibits the proteolytic cleavage of Plasmodium  
 CC MSP-1\_4\_2, compared to natural MSP-1\_1\_9. The non-natural variants of the  
 CC present invention are useful for immunising a mammal against malaria, and  
 CC can be used to treat malaria. The present sequence is the MSP-119  
 CC coding sequence.  
 XX  
 SQ Sequence 333 BP; 107 A; 75 C; 73 G; 78 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 1.95e-35 Length: 333  
 Score: 525.00 Matches: 92  
 Percent Similarity: 93.20% Conservative: 4  
 Best Local Similarity: 89.32% Mismatches: 3  
 Query Match: 23.08% Indels: 4  
 DB: 22 Gaps: 1  
 US-10-057-531A-2 (1-431) x AAC68977 (1-333)  
 Qy 329 IleGluGlyLysPheGlnAspMetLeuAsnIleSerGlnHisGlnCysValLysGln 348  
 Db 22 ATTGAAGGTAGACAC-----AACATTGCCCAACACCAATCGGTAAAGACAA 69  
 Qy 349 CysProGluAsnSerGlyCysPheArgHisLeuAspGluArgGluGluCysLysCysLeu 368  
 Db 70 TGTCCACAAACTCCGATGTTTCAGACATCTGCACGAGAGAGAAGATGTAGTGTCTG 129  
 Qy 369 LeuAsnTyrLysGlnGluGlyAspLysCysValGluAsnProAsnProThrCysAsnGlu 388  
 Db 130 TTGAACCTACAGCAGAGAGGTGATAAGTGTGTGAGAACCCAAACCCCTACCTGTAAACGAG 189  
 Qy 389 AsnAsnGlyGlyCysAspAlaAspAlaThrCysThrGluGluAspSerGlySerArg 408  
 Db 190 AACACCGTGGATCGGACGCTAGCTAAGTGCACCAAGAGACTCTGGTCTTAAACGGA 249  
 Qy 409 LysLysIleThrCysGluCysThrLysProAspSerTyrProLeuPheAspGlyIlePhe 428  
 Db 250 AAGAAGATTACTTGGGAATGTACTTAAGCAGACTCTTACCCCTTTGTCGATGAATCTTC 309  
 Qy 429 CysSerSer 431  
 Db 310 TGTCTCTCC 318  
 RESULT 22  
 AAD22459  
 ID AAD22459 standard; DNA; 309 BP.  
 XX  
 AC AAD22459;  
 XX  
 DT 12-FEB-2002 (first entry)



```
XX Plasmodium falciparum PfMSP1.19 Insert in pUC105-01 vector.
DE PUC105-01 vector; Immunostimulant; vaccine; immunisation; therapeutic;
KW C3d; Immune response; ds.
XX Plasmodium falciparum.
OS WO200177324-A1.
XX 18-OCT-2001.
PN 09-APR-2001; 2001WO-GB01599.
PD 08-APR-2000; 2000GB-0008582.
XX (ADPR-) ADPROTECH LTD.
XX Steward M, Cox VF;
PI WPI; 2002-010909/01.
XX Novel variant DNA sequence useful in DNA vaccine, encodes naturally
PT occurring protein and comprises a sequence non-identical to naturally
PT occurring DNA sequence encoding the protein -
PS Example 13; Page 65; 87pp; English.
XX The invention relates to a variant DNA sequence useful in DNA vaccines.
CC The DNA sequence encodes a naturally occurring protein such as C3d which,
CC by virtue of third base redundancy and other variations permissible
CC within an amino acid codon, is non-identical to the naturally occurring
CC DNA sequence encoding that protein. The DNA sequence is useful in a DNA
CC immunisation vector to encode one or more naturally occurring human or
CC non-human proteins with immunomodulatory properties. The DNA sequence
CC is useful for inducing an immune response to an antigen in a human or
CC animal. A pharmaceutical composition is useful for introducing a DNA
CC sequence encoding a naturally occurring protein into a human or animal,
CC by administering the pharmaceutical composition into the human or animal,
CC where the administration results in a therapeutic effect on the human or
CC animal. The present sequence is Plasmodium falciparum PfMSP1.19 insert
CC in pUC105-01 vector
XX SQ Sequence 309 BP; 94 A; 71 C; 69 G; 75 T; 0 other;

Alignment Scores:
Pred. No.: 309 Length: 309
Score: 521.00 Matches: 89
Percent Similarity: 97.8% Conservative: 3
Best Local Similarity: 94.68% Mismatches: 2
Query Match: 22.90% Indels: 0
DB: 24 Gaps: 0

US-10-057-531A-2 (1-431) x AAD22459 (1-309)
QY 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357
Db 7 AACATTGCCCAACCAATCGCTTAAGAAGCAATGCCACAAATCCCGATGTTTCAGA 66
QY 358 HisLeuAspGluArgGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377
Db 67 CATCTGGACGAGAGAGAAGAAATGTAAGTCTCTGTTGAATCAACGAGGAAAGCTGATAAG 126
QY 378 CysValGluAsnProAsnProThrCysAsnGluAsnAsnGlyCysAspAlaAspAla 397
Db 127 TGTGTTGAGACCAACCAACCTACCTGTACGAGACACGGTGGATGCCACGCTGACGCT 186
QY 398 ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417
Db 187 AAGTGCACCGAAGAGACTCTGTTCTTAACGGAAGAAAGATTACTTGCGAATCTACTAAG 246
QY 418 ProAspSerTyrProLeupheAspGlyIlePheCysSerSer 431
|||||
```

```
Db 247 CCAGACTCTTACCCTTTGTTGATGATGAATCTTCTGTCTTCC 288
RESULT 23
AAD22460
ID AAD22460 standard; DNA; 3147 BP.
XX AC AAD22460;
XX 12-FEB-2002 (first entry)
DE PfMSP1.19-human C3d3 DNA coding sequence from pVK104-01 vector.
XX Human; immunostimulant; vaccine; immunisation; therapeutic; C3d;
KW immune response; pVK104-01 vector; ds.
XX Chimeric - Homo sapiens.
OS Chimeric - Plasmodium falciparum.
OS Chimeric - Unidentified.
XX FH Key Location/Qualifiers
FT CDS 1..3147
FT /tag= a
FT /product= "Human C3d3-PfMSP1.19 protein"
FT /note= "CDS does not include stop codon"
FT /partial
FT sig_peptide 1..72
FT /tag= b
FT mat_peptide 73..3147
FT /tag= c
FT /product= "Mature human C3d3 protein"
FT /note= "Antigen sequence"
FT /tag= d
WO200177324-A1.
XX 18-OCT-2001.
XX 09-APR-2001; 2001WO-GB01599.
XX 08-APR-2000; 2000GB-0008582.
XX (ADPR-) ADPROTECH LTD.
XX Steward M, Cox VF;
XX WPI; 2002-010909/01.
XX Novel variant DNA sequence useful in DNA vaccine, encodes naturally
XX occurring protein and comprises a sequence non-identical to naturally
XX occurring DNA sequence encoding the protein -
XX Claim 16; Page 66-68; 87pp; English.
XX The invention relates to a variant DNA sequence useful in DNA vaccines.
XX The DNA sequence encodes a naturally occurring protein such as C3d which,
XX by virtue of third base redundancy and other variations permissible
XX within an amino acid codon, is non-identical to the naturally occurring
XX DNA sequence encoding that protein. The DNA sequence is useful in a DNA
XX immunisation vector to encode one or more naturally occurring human or
XX non-human proteins with immunomodulatory properties. The DNA sequence
XX is useful for inducing an immune response to an antigen in a human or
XX animal. A pharmaceutical composition is useful for introducing a DNA
XX sequence encoding a naturally occurring protein into a human or animal,
XX by administering the pharmaceutical composition into the human or animal,
XX where the administration results in a therapeutic effect on the human or
XX animal. The present sequence is Plasmodium falciparum PfMSP1.19-human
XX C3d3 DNA coding sequence from vector pVK104-01.
XX SQ Sequence 3147 BP; 869 A; 754 C; 834 G; 690 T; 0 other;

Alignment Scores:
Pred. No.: 6,41e-34 Length: 3147
```

Score: 521.00 Matches: 89  
Percent Similarity: 97.87% Conservative: 3  
Best Local Similarity: 94.68% Mismatches: 2  
Query Match: 22.90% Indels: 0  
DB: 24 Gaps: 0

US-10-057-531A-2 (1-431) x AAD22460 (1-3147)

Qy 338 AsnIleSerGlnHisGlnCysValIysLysGlnCysProGluAsnSerGlyCysPheArg 357  
Db 73 AACATTCGCCAACACCACTCGTTAAGAGCAATGTCCACAAACTCCGGATGTTTCAGA 132  
Qy 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377  
Db 133 CATCTGCACGAGAGAGAGAAATGTAAGTGTCTGTGAACACAAAGAGAGGTGATAAG 192  
Qy 378 CysValGluAsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAla 397  
Db 193 TGTCTTGAGAACCAACCCCTACTGTACGAGAACACCGTGGATCGACGCTGACGCT 252  
Qy 398 ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417  
Db 253 AAGTGCACCGAAGAAAGACTCTGGTTCTTAACGGAAGAAGATTACTTCGGAATGTACTAAG 312  
Qy 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431  
Db 313 CCAGACTCTTACCCTTTGTTTCGATGGAATCTCTGTTCTTCC 354

RESULT 24

AAD22461

ID AAD22461 standard; DNA: 3147 BP.

XX AAD22461;

AC AAD22461;

DT 12-FEB-2002 (first entry)

DE Human C3d3 DNA-PfMSPl.19 coding sequence from pVK104-02 vector.

KW Human; Immunostimulant; vaccine; immunisation; therapeutic; C3d;  
KW Immune response; pVK104-02 vector; ds.

OS Chimeric - Homo sapiens.

OS Chimeric - Plasmodium falciparum.

OS Chimeric - Unidentified.

FX Key Location/Qualifiers

FT CDS 1..3147

FT /\*tag= a

FT /product= "Human C3d3-PfMSPl.19 protein"

FT /note= "CDS does not include stop codon"

FT /partial

FT sig\_peptide 1..72

FT /\*tag= b

FT mat\_peptide 73..3147

FT /\*tag= c

FT /product= "Mature human C3d3-PfMSPl.19 protein"

FT 2845..3142

FT /\*tag= d

FT /note= "Antigen sequence"

XX WO200177324-A1.

XX 18-OCT-2001.

XX 09-APR-2001; 2001WO-GB01599.

XX 08-APR-2000; 2000GB-0008582.

XX (ADPR-) ADPROTECH LTD.

XX Steward M, Cox VF;

XX WPI; 2002-010909/01.

XX

PT Novel variant DNA sequence useful in DNA vaccine, encodes naturally occurring protein and comprises a sequence non-identical to naturally occurring DNA sequence encoding the protein

PS Claim 16: Page 69-71; 87pp; English.

XX The invention relates to a variant DNA sequence useful in DNA vaccines.

CC The DNA sequence encodes a naturally occurring protein such as C3d which, by virtue of third base redundancy and other variations permissible within an amino acid codon, is non-identical to the naturally occurring DNA sequence encoding that protein. The DNA sequence is useful in a DNA immunisation vector to encode one or more naturally occurring human or non-human proteins with immunomodulatory properties. The DNA sequence is useful for inducing an immune response to an antigen in a human or animal. A pharmaceutical composition is useful for introducing a DNA sequence encoding a naturally occurring protein into a human or animal by administering the pharmaceutical composition into the human or animal, where the administration results in a therapeutic effect on the human or animal. The present sequence is Plasmodium falciparum pfMSPl.19-human C3d3 DNA coding sequence from vector pVK104-02.

XX Sequence 3147 BP; 869 A; 754 C; 834 G; 690 T; 0 other;

SQ Alignment Scores:

Pred. No.: 6,41e-34 Length: 3147  
Score: 521.00 Matches: 89  
Percent Similarity: 97.87% Conservative: 3  
Best Local Similarity: 94.68% Mismatches: 2  
Query Match: 22.90% Indels: 0  
DB: 24 Gaps: 0

US-10-057-531A-2 (1-431) x AAD22461 (1-3147)

Qy 338 AsnIleSerGlnHisGlnCysValIysLysGlnCysProGluAsnSerGlyCysPheArg 357

Db 2845 AACATTCGCCAACACCACTCGTTAAGAGCAATGTCCACAAACTCCGGATGTTTCAGA 2904

Qy 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377

Db 2905 CATCTGCACGAGAGAGAAATGTAAGTGTCTGTGAACACAAAGAGGTGATAAG 2964

Qy 378 CysValGluAsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAla 397

Db 2965 TGTCTTGAGAACCCAAACCCCTACTGTAAACGAGAACACCGTGGATCGACGCT 3024

Qy 398 ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417

Db 3025 AAGTGCACCGAAGAGACTCTGGTTCTACGGAAGAAGATTACTTCGGAATGTACTAAG 3084

Qy 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431

Db 3085 CCAGACTCTTACCCTTTGTTTCGATGGAATCTCTGTTCTTCC 3126

RESULT 25

AAD22462

ID AAD22462 standard; DNA: 309 BP.

XX AAD22462;

AC AAD22462;

XX 12-FEB-2002 (first entry)

DE Plasmodium falciparum MSPl.19 (pfMSPl.19) mutant DNA.

XX Immunostimulant; vaccine; immunisation; therapeutic; immune response;

KW MSPl.19 gene; mutant; ds.

XX Plasmodium falciparum.

OS Synthetic.

XX Key Location/Qualifiers

FT misc\_feature 40..42

FT /\*tag= a









CC multistage vaccine against P. vivax.

XX Sequence 6101 BP; 2077 A; 1168 C; 1534 G; 1322 T; 0 other;

#### Alignment Scores:

Pred. No.: 8.57e-14 Length: 6101  
Score: 284.50 Matches: 85  
Percent Similarity: 40.00% Conservativity: 55  
Best Local Similarity: 24.29% Mismatches: 106  
Query Match: 12.51% Indels: 104  
DB: 22 Gaps: 10

US-10-057-531A-2 (1-431) x AAH47055 (1-6101)

QY 101 AsnLeuAsnLeuAspIleLeuAsnSerArgLeuLysLysArgLysTyrPheLeuAsp 120  
DB 5295 AATTTGGACAGCAATGCTACTCAGTTAAAGAAAGATTAAAGGGCAATTTATATGGAT 5354  
QY 121 Val-----LeuGluSerAspLeuMetGlnPheLysHisIleSe 133  
DB 5355 TTGTAATTAATTTGCGGTAATATAGACCGCAGATATACAGATGGATTTCGAGAATG 5414  
QY 133 rSerAsnGluTyrIleIleGluAspSerPheLysLysLeuAsnSerGluGluLysAsnTh 153  
DB 5415 GGGAGGGGATTACGTGTCAGAA-----TTGCCACAGAGTGCAGAAA----- 5457  
QY 153 rLeuLeuLysSerTyrLysTyrIleLysGluSerValGluAsnAspIleLysPhe----- 171  
DB 5458 -----CTGAAGAGAAATGTGATGAGAAATCAATTAATACTGA 5495  
QY 172 -----AlaGlnGluGlyIleSerTyrTyrGluLysValLe 183  
DB 5496 TAAAAAGTATGTAAGGTACCACCATGTCATAATCGCTGTAATCATATGATCAATGGAT 5555  
QY 183 uAlaLysTyrLysAsp-----AspLeuGluSerIleLysLysValIleLysGluLuluy 201  
DB 5556 ACCGAGAAAANAATCAATGGGATGCTGTCATAAATAATCAATGAAGTGAAGAAACGC 5615  
QY 201 sGluLysPheProSerProProThrThrProProSerProAlaLysThrAspGluGl 221  
DB 5616 AGAAAAGGTTACAGCGCAGGTATCGTAACCTTATGATATATAAATACAGAGGTAGA 5675  
QY 221 nLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeuLysAsnLeu 241  
DB 5676 TGAATTTAAC-----GAGGTGGCTTTTGAGAATGA 5705  
QY 241 uValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsnVa 261  
DB 5706 AATTACAACGCTGATGCTATATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5753  
QY 261 lGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLysAlaIleAspAs 281  
DB 5754 TGAAGAG-----CCTAAAAAATACTCAGGAAGTTCTGCACAA 5792  
QY 281 pLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLysLeuIleAsnAs 301  
DB 5793 TGTGGACAATGCTGCTAAATCTCCATGGACTATGAGC----- 5829  
QY 301 pasphThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuValGlnAsnPhePr 321  
DB 5829 ----- 5829  
QY 321 oAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeuAsnIleSerGl 341  
DB 5830 -----TCCGA 5834  
QY 341 nHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArgHisLeuAspGl 361  
DB 5835 GCACACATGATAGACACCAATGTGCTGATATGACCGCTCTAGTACTTGGACGG 5894  
QY 361 uArgGluGluCysLysCysLysLeuAsnTyrLysGlnGluGlyAspLysCysValGluAs 381  
DB 5895 AACGGAAGATGGATGCTTGTACCTTTAAAGAAAGGCGGCAAGTGTGCTGCCAGC 5954

QY 381 nProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAlaThrCysThrGl 401  
DB 5955 ATCGAATGTGACTTGTAGGATAACAAATGGTGTGTCGCCCTGAAGCTGAATGTAAAT 6014  
QY 401 uGluAspSerGlySerSerArgLysLysIleThrCysGluCysThrLysProAspSerTy 421  
DB 6015 GACGGACAGC-----AATAAATCGTCTCTAAATGTACTAAAGAGGTTCTGA 6062  
QY 421 rProLeuPheAspGlyIlePheCysSer 430  
DB 6063 GCCACCCCTTTGAGGGAGTTTCTGTAGC 6090

#### RESULT 31

AAH47056  
ID AAH47056 standard; DNA; 5643 BP.  
XX  
AC AAH47056;  
XX  
DT 29-OCT-2001 (first entry)  
XX  
DE Synthetic gene ViVac2bias nucleotide sequence.  
XX  
KW Multivalent protein; immune response; Plasmodium vivax; parasite;  
KW protozoacide; vaccine; malaria; recombinant; ViVacip; ds.  
XX  
OS Synthetic.  
OS Plasmodium vivax.  
PN WO20015181-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 29-JAN-2001; 2001WO-US02937.  
XX  
PR 31-JAN-2000; 2000US-0179213.  
XX  
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.  
PI Lal AA, Xiao L, Zhou Z;  
DR WPI: 2001-514557/56.  
DR P-PSDB; AAB85698.

New recombinant multivalent protein comprising antigenic determinants derived from more than one stage in a life cycle of Plasmodium vivax, useful as a vaccine for treating, preventing and reducing malarial infection

Claim 9; Page 56-59; 59pp; English.

The invention relates to recombinant multivalent proteins (I) that stimulate an immune response to Plasmodium vivax. (I) comprises antigenic determinants, fragments or conservative substitutions, derived from more than one stage in a life cycle of a Plasmodium vivax parasite. (I) is useful as a vaccine for stimulating an immune response, specifically a protective immune response that confers increased resistance to infection by Plasmodium parasites, such as P. vivax. (I) is especially useful in the treatment, prevention and reduction of malarial infection, as research or diagnostic reagents for the detection of Plasmodium species in a biological sample, and for conferring immunity against multiple stages of the malarial parasite. The antibodies produced are useful for the detection or measurement of antigenic epitopes derived from one or more stages in a life cycle of a parasite, particularly P. vivax. The vaccine comprising the recombinant proteins, is cost-effective, health-promoting intervention for controlling, preventing or treating the incidence of malaria. The present sequence represents the nucleotide sequence of a synthetic gene ViVac2bias.

Sequence 5643 BP; 1584 A; 1197 C; 1503 G; 1359 T; 0 other;

Alignment Scores:

Pred. No.: 9.45e-14 Length: 5643

Score:	283.50	Matches:	96
Percent Similarity:	37.44%	Conservative:	59
Best Local Similarity:	23.19%	Mismatches:	129
Query Match:	12.46%	Indels:	131
DB:	22	Gaps:	13
US-10-057-531A-2 (1-431) x AAH47056 (1-5643)			
Qy	32	AspSerProAspLeuGlyThr-----AspAspAspAspLysAlaMetAlaAspIle	48
Db	4742	GATACTCAAGTTGTGGAGAACAACTTGAGATCCATCTCCGAACATGATGAGAGGCTC	4801
Qy	49	GlySerIleGluGlyArgGlyThrMetAlaIleSerValThrMetAspAsnIleLeuSer	68
Db	4802	AACAGCGTAGAACGACGTGGGACGAGTCTAAGG-----	4837
Qy	69	GlyPheGluAsnGluTyrAspValIleTyrLeuLysProLeuAlaGlyValTyrArgSer	88
Db	4838	-----CTCAAACTCG- ACTGCTATGATGATCTACC	4866
Qy	89	LeuLysLysGlnIleGluLysAsnIlePheThrPheAsnLeuAsnLeuAsnAspIleLeu	108
Db	4867	GTTAAGAGAGATTGAAGGGTAAC---TTCACTGGATCTGTAAGTTGAACGTTGCTGTT	4923
Qy	109	AsnSerArgLeuLysLysArgLysTyrPheLeuAspValLeuGluSerAspLeuMetGln	128
Db	4924	AAC-----ATCGAGCCCATGATCTACAGA	4947
Qy	129	PheLysHisIleSerSerAsnGluTyrIleIleGluAspSerPheLysLeuLeuAsnSer	148
Db	4948	TGGATCAGAGAGTGGGAAGAGATTAGCTGCCGAG-----TTGCCCACTGAG	4995
Qy	149	GluGlnLysAsnThrLeuLeuLysSerTyrLysTyrIleLysGluSerValGluAsnAsp	168
Db	4996	GTGCAAAAG-----CTGAAGGAGAGTGTGATGGAAG	5028
Qy	169	IleLysPhe-----AlaGlnGluGlyIleSerTyr	178
Db	5029	ATCAACTACACTGATAAGAGGTTTGAAGTTCACCATGTCACAAACGGTGTAAAGTCC	5088
Qy	179	TyrGluLysValLeuAlaLysTyrLysAsp-----AspLeuGluSerIleLysLysVal	196
Db	5089	TACGATCAATGGATCACTAGAGAACAAACAAATGGGATGTTCTGTCCAAACAAAGTTTCA	5148
Qy	197	IleLysGluGluLysGluLysPheProSerSerProThrThrProProSerProAla	216
Db	5149	TCTGTAAGAACGCTGAGAGGTTTCAGACTGCTGATCGTTTACTCCATACGATACCTG	5208
Qy	217	LysThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThr	236
Db	5209	AAGCAGAGTTGGATGAGTTCAC-----GAGGTG	5238
Qy	237	LeuTyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIle	256
Db	5239	GCTTTCGAGACGAGATCAACACCGTGTGCTTATATATCGAGTTGTGTGTT-----	5292
Qy	257	AsnAspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeu	276
Db	5293	-----TGTTCCGTTGAGGAG-----GCTAAGAGAACACTCAG	5325
Qy	277	LysAlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLys	296
Db	5326	GAGGTTGTGACTACGTGGACACGCTGCTAAGTCTCCATGAGTATGCTC-----	5376
Qy	297	LysLeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeu	316
Db	5376	-----	5376
Qy	317	ValGlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMet	336
Db	5376	-----	5376
Qy	337	LeuAsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPhe	356



CC animal. The present sequence is Plasmodium yoelii MSP1 (PymSP1.19)  
 CC antigen sequence and murine C3d chimeric coding sequence from pVK96-02  
 CC vector.

XX  
 SQ Sequence 3252 BP; 864 A; 786 C; 858 G; 744 T; 0 other:

## Alignment Scores:

Pred. No.: 5.88e-14 Length: 3252  
 Score: 282.50 Matches: 119  
 Percent Similarity: 39.76% Conservative: 83  
 Best Local Similarity: 23.43% Mismatches: 175  
 Query Match: 12.42% Indels: 131  
 DB: 24 Gaps: 21

US-10-057-531A-2 (1-431) x AAD22449 (1-3252)

QY 22 AlaAlaAlaLysPheGluArgGlnHisMetAspSerProAspLeuGlyThrAspAsp 41  
 DB 1819 GCGCTGGCGCAGTACCAGACGACGTGCCAGATCAACAAGACCTCAATATGACGCTCAGT 1878  
 QY 42 AspLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMetAlaIleSerVal 61  
 DB 1879 TTTCAC---TTGCCATCATCCGGAGC-----GCGGAGGTGGGAGCGGAGGGCGGT 1929  
 QY 62 ThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIleTyrLeuLysPro 81  
 DB 1930 ACCTCGGATCTACCCCGCAGGC---TCTGGGAACACAGACATGATGGCATGACACCA 1986  
 QY 82 ----LeuAlaGlyValTyrArgSerLeuLysGlnIleGluLysAsnIlePheThr 99  
 DB 1987 ACAGTCATTGGCTGCTACACTACCTGGACACCGACAGCTGGGAGAG-----2034  
 QY 100 PheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyrPheLeu 119  
 DB 2035 -----TTCGCATAGACAGAGGCAAGGCCCTG 2064  
 QY 120 AspValLeuGluSerAspLeuMetGln-----PheLysHisIleSerSerAsnGlu 136  
 DB 2065 GAGCTCATCAAGAAAGGGTACACCCAGCAGCTGGCCCTTCAACAGCCAGCTCT-----2118  
 QY 137 TyrIleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLys 156  
 DB 2119 -----GCCTATGCTGCCTTCAACAACAGCCGCCCGCCAGCACCTGGCTGACA 2163  
 QY 157 SerTyrLys-----159  
 DB 2164 GCCTACGTGGTCAAGGTCTTCTCTAGCTGCCAACCTCATGCCATCGACTCTCACGTC 2223  
 QY 160 -----TyrIleLysGluSerValGluAsnAspIleLysPheAla 172  
 DB 2224 CTGTGTGGGCTGTTAAATGGTTGATCTGGAGAAACAGACGCGGATGGTGTCTTTTCA 2283  
 QY 173 GlnGluGlyIleSerTyrTyrGluLysValLeuAlaLysTyrLysAspLeuGluSer 192  
 DB 2284 GAGGATGGCCCGCGTATTCACCAAGAAATGATTGGTGGCTCCGGAACCCCAAGGAGGCA 2343  
 QY 193 -----IleLysLysValIleLysGluGluLysGluLysPhePro 205  
 DB 2344 GATGTCTCACTCACACCCTTCGCTCTCATCGCAGTCCGAGAACCCAGGACATCTGTGAG 2403  
 QY 206 SerSerProProThrProProSerProAlaLysThrAspGlnLysLysGluSer 225  
 DB 2404 GGGCAGGTCAATAGCTTCTGGGAGCATCAACAGGCGGAGGTATATTGAAGCCAGT 2463  
 QY 226 LysPhe-----LeuProPheLeuThrAsnIleGluThrLeuTyrAsnLeuVal 242  
 DB 2464 TCATGAACCTGCAGACCATACAGATGGCCATTGCTGGGTATGCCCTGGCCCTGATG 2523  
 QY 243 AsnLysIleAspAsp----TyrIleLeuAsnLeuLysAlaLysIleAsnAspCysAsn---260  
 DB 2524 ACAAACTGGAGAACCTTACCTCGCAAGTTCCTGCAACACAGCCCAAGATCGGAACCGC 2583  
 QY 261 ----ValGluLysAspGluAla-----HisValLysIleThrLysLeuSerAspLeuLys 277

DB 2584 TGGGAGGCGCTGACCAGCAGCTCTACAACGTAGAGCCACATCCCTACGCCCTCTGGCC 2643  
 QY 278 AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys 297  
 DB 2644 CTGCTG-----CTGCTGAAA-----GACITTGACTCTGTGCCCT 2679  
 QY 298 LeuIle-----299  
 DB 2680 GTAGTCGCTGGCTCAATGAGCANAAGATACTACGGAGCGGCTATGGCTCCACCCAGGCT 2739  
 QY 300 -----AsnAspAsp 302  
 DB 2740 ACCTTCATGGTATTCGAAGCCTTGGCCCAATATCAACACAGATGTCCTCGACCATAGGAC 2799  
 QY 303 ThrLysLysAspMet-----307  
 DB 2800 TTGAACATGGATGTGTCTCCACCTCCCGCAGCAGTGGATCTCACATTTGGCTCTATTGCT 2859  
 QY 308 LeuGlyLysLeuLeuSerThrGlyLeuVal---GlnAsnPheProAsnThrIleIleSer 326  
 DB 2860 TTGAACAACCTTGAACAAGTCTGGTTGGTAGGAGAGGTGAGTCTAAGAAGATTTTGGCT 2919  
 QY 327 LysLeuIleGluGlyLysPheGlnAspMetLeuAsnIleSer---GlnHisGlnCysVal 345  
 DB 2920 AAGATGCTGAACATGGACGCTATGGACTTGTGGTGTTCACCTAAGCATGTTGTGTT 2979  
 QY 345 ---LysLysGlnCysProGluAsnSerGlyCysPheArgHisLeuAspGluArgGluGlu 364  
 DB 2980 GACACTAGAGACATCTTAAGACGCTGGATGTTTCAGAGCAGCAACCGTACTGAAGAG 3039  
 QY 365 CysLysCysLeuLeuAsnTyrLysGln---GluGlyAspLysCysValGluAsnProAsn 383  
 DB 3040 TGGAGATGTTTGGTTTACAAGAAGGTGAGGTAAACCTGCGTTGAGAACAACAAC 3099  
 QY 384 ProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAlaThrCysThrGluGluAsp 403  
 DB 3100 CCTACTTGGGACATCAACAACGCTGGATGTGACCAACGCGCTTGTCTCAAAACGCTGAA 3159  
 QY 404 SerGlySerSerArgLysLysIleThrCysGluCysThrLysProAspSerTyrProLeu 423  
 DB 3160 TCTACCGAAATCCCAAGAAGATTATTGCACCTGTAAAGAACCAACCCCTAACGCCCTAC 3219  
 QY 424 PheAspGlyIlePheCysSerSer 431  
 DB 3220 TAGGAGGTGTTTCTGTTCTTCT 3243  
 RESULT 33  
 AAD21371  
 ID AAD21371 standard; DNA; 870 BP.  
 XX  
 AC AAD21371;  
 XX  
 DT 28-JAN-2002 (first entry)  
 XX  
 DE Helicobacter pylori His-HOP38(-11) polypeptide encoding DNA.  
 KW Polypeptide delivery system; immune stimulating complex; ISCOM;  
 KW cholesterol; saponin; phospholipid; medicament; vaccine; therapy;  
 XX Helicobacter pylori infection; HOP38; anti-bacterial; ds.  
 OS Helicobacter pylori.  
 FH Key Location/Qualifiers  
 CDS 1..870  
 FT /tag= a  
 FT /product= "Helicobacter pylori His-HOP38(-11)  
 FT polypeptide"  
 PN WO200176625-A1.  
 XX 18-OCT-2001.  
 PD  
 XX

PF 09-APR-2001; 2001WO-SE00799.

XX 12-APR-2000; 2000GB-0008879.

PR (ASTR ) ASTRAZENECA AB.

PA Shaprio A, Sanyal G;

XX WPI; 2001-663016/76.

XX P-PSDB; AAEL13033.

XX Producing a polypeptide delivery system useful in a vaccine to treat

PT infection by mixing together the polypeptide, cholesterol, saponin, and

PT a phospholipid in presence of a nonionic detergent and a second

PT detergent

XX Disclosure; Page 38-39; 43pp; English.

XX The invention relates to producing a polypeptide delivery system

CC comprising an immune stimulating complex (ISCOM) coupled to a

CC polypeptide. The method involves mixing the polypeptide, cholesterol,

CC saponin and a phospholipid in the presence of a nonionic detergent

CC and a second detergent to form a solution and removing the detergent

CC from the mixture to form the ISCOM. The polypeptide delivery system

CC is used as a medicament for the manufacture of a vaccine for

CC administration to mammalian patients, to treat and prevent

CC Helicobacter pylori infection. The present sequence is a DNA encoding

CC His-HOP38(-11), a Helicobacter pylori polypeptide used in the

CC polypeptide delivery system.

XX SQ Sequence 870 BP; 235 A; 170 C; 214 G; 251 T; 0 other;

Alignment Scores:

Pred. No.: 2.57e-14 Length: 870

Score: 278.50 Matches: 53

Percent Similarity: 96.36% Conservative: 0

Best Local Similarity: 96.36% Mismatches: 1

Query Match: 12.24% Indels: 1

DB: 22 Gaps: 1

US-10-057-531A-2 (1-431) x AAD21371 (1-870)

QY 1 MethHisHisHisHisHisHisSerSerGlyLeuValProArgGlySerGlyMetLysGlu 20

Db 1 ATGCACCATCATCATCATCATCTCTGGTCTGGTCCACCGCGTCTGGTATGAAAGAA 60

QY 21 ThrAlaAlaAlaLysPheGluArgGlnHisMetAspSerProAspLeuGlyThrAspAsp 40

Db 61 ACCGCTGCTGCTAAATTCGAACGCCAGCACATGCACAGCCAGATCTGGGTACCGAGCAG 120

QY 41 AspAspLysAlaMetAlaAspIleGlySerIleGluGlyArgGly 55

Db 121 GACGACAAAGGCCATGGCTGATATCGGATCC---GAAGGTGATGTT 162

RESULT 34

AAH43782

ID AAH43782 standard; cDNA; 870 BP.

XX AAH43782;

XX 30-JAN-2002 (first entry)

DE HOP38 coding sequence #5.

XX H. pylori; HOP38; polypeptide delivery system; polar head group;

XX immune stimulating complex; ISCOM; antigenic fragment; saponin;

XX cholesterol; phospholipid; detergent; ss.

XX Helicobacter pylori.

OS WO200176623-A1.

PN 18-OCT-2001.

PD

XX

PF 09-APR-2001; 2001WO-SE00800.

XX 12-APR-2000; 2000GB-0008877.

XX (ASTR ) ASTRAZENECA AB.

XX Shaprio A, Sanyal G;

XX WPI; 2002-025884/03.

XX P-PSDB; AAB47732.

XX Production of a polypeptide delivery system useful as a medicament

PT comprises mixing together the polypeptide, cholesterol, saponin, and a

PT phospholipid with a polar head group, in the presence of a detergent

XX Disclosure; Page 45; 48pp; English.

XX The sequences given in AAH43778-82 encode H. pylori HOP38 proteins

CC which were used in the method of the invention. The method of the

CC invention is a process for production of a polypeptide delivery system

CC comprising an immune-stimulating complex (ISCOM) coupled to a

CC polypeptide of H. pylori or its antigenic fragment. The method comprises:

CC mixing the polypeptide, cholesterol, a saponin, and a phospholipid

CC having a polar head group; and removing the detergent from the mixture

CC to form an ISCOM. The method has a broad applicability to polypeptides,

CC including polypeptides that are unsuitable to prior art processes.

XX SQ Sequence 870 BP; 235 A; 170 C; 214 G; 251 T; 0 other;

Alignment Scores:

Pred. No.: 2.57e-14 Length: 870

Score: 278.50 Matches: 53

Percent Similarity: 96.36% Conservative: 0

Best Local Similarity: 96.36% Mismatches: 1

Query Match: 12.24% Indels: 1

DB: 22 Gaps: 1

US-10-057-531A-2 (1-431) x AAH43782 (1-870)

QY 1 MethHisHisHisHisHisHisSerSerGlyLeuValProArgGlySerGlyMetLysGlu 20

Db 1 ATGCACCATCATCATCATCATCTCTGGTCTGGTCCACCGCGTCTGGTATGAAAGAA 60

QY 21 ThrAlaAlaAlaLysPheGluArgGlnHisMetAspSerProAspLeuGlyThrAspAsp 40

Db 61 ACCGCTGCTGCTAAATTCGAACGCCAGCACATGCACAGCCAGATCTGGGTACCGAGCAG 120

QY 41 AspAspLysAlaMetAlaAspIleGlySerIleGluGlyArgGly 55

Db 121 GACGACAAAGGCCATGGCTGATATCGGATCC---GAAGGTGATGTT 162

RESULT 35

ABA95618

ID ABA95618 standard; cDNA; 2472 BP.

XX ABA95618;

XX 21-MAR-2002 (first entry)

DE Thioresoxin/deoxyxylulose 5-phosphate synthase fusion protein cDNA.

XX DXPS; deoxyxylulose 5-phosphate synthase; pyruvate; thioresoxin;

XX glyceraldehyde 3-phosphate; plant growth modulator;

XX microbial growth modulator; enzyme; gene; ss.

XX Chimeric - Arabidopsis thaliana.

OS Chimeric - Synthetic.

XX Key Location/Qualifiers

XX CDS 1..2472

XX /\*tag="a

XX /partial

/product="trxA/txPS fusion protein"  
/note="No stop codon given"

US6326164-B1.

04-DEC-2001.

27-JUL-2000; 2000US-0626589.

27-JUL-2000; 2000US-0626589.

(PARA-) PARADIGM GENETICS INC.

Rice JW, Kloti AS, Crawford JM, Lanning B, Stewart SJ;

WPI; 2002-121106/16.

P-PSDB; AM48247.

Assay for determining deoxyxylulose 5-phosphate synthase activity by measuring pyruvate depletion, useful for screening compounds that inhibit or enhance this activity which is useful for modulating plant and microbial growth.

Disclosure: Columns 27-30; 26pp; English.

The present sequence is the coding sequence for a fusion protein comprising a truncated deoxyxylulose 5-phosphate synthase (DXPS) from *Arabidopsis thaliana* and thioredoxin (trxA) from vector pET32. The N-terminal 58 amino acids of the DXPS protein were removed to generate the truncated protein. The truncated DXPS protein possesses DXPS activity. The full-length DXPS is given in AM48245. The present invention relates to a method for determining DXPS activity. The method comprises contacting pyruvate, and optionally glyceraldehyde 3-phosphate with DXPS and then determining the concentration of remaining pyruvate and/or glyceraldehyde 3-phosphate. The method is useful for screening for inhibitors and enhancers of DXPS activity which will have use as modulators of plant and microbial growth.

Sequence 2472 BP; 662 A; 517 C; 640 G; 653 T; 0 other;

#### Alignment Scores:

Pred. No.:	9, 1e-14	Length:	2472
Score:	278.50	Matches:	76
Percent Similarity:	49.75%	Conservative:	25
Best Local Similarity:	37.44%	Mismatches:	49
Query Match:	12.24%	Indels:	53
DB:	24	Gaps:	7

US-10-057-531A-2 (1-431) x ABA95618 (1-2472)

QY	1	MethHisHisHisHisSerSerGlyLeuValProArgGlySerGlyMetLysGlu	20
DB	346	ATGCACCATCATCATCTCTCTGCTGCTGCCACGCGGTCTCGTATGAAGAA	405
QY	21	ThrAlaAlaLysPheGluArgGlnHisMetAspSerProAspLeuGlyThrAspAsp	40
DB	406	ACCGTGTCTGCTAAATTCGAAGCCACGACATGGACAGCCAGATCTGGTACCGAC	465
QY	41	AspAspLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMetAlaIleSer	60
DB	466	GAGCAAGGCCATGGCTGATATCGGATCCGCTTCA	501
QY	61	ValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIleTyrLeuLys	80
DB	501	-----	501
QY	81	ProLeuAla-----GlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIlePhe	98
DB	502	---CTTCGACAGAGGGTGAATATTATTCAC---AACAGACCACCACTCCATTACTTGAC	555
QY	99	ThrPheAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyrPhe	118
DB	556	ACTATTAACTACCAATCCAC---ATGAAATAATCTTTCTGTCGAAGGAAGTGAACAACTT	612

QY	119	LeuAspValLeuGluSerAspLeuMetGlnPheLys-----HisIle	132
DB	613	TCTGATGAGCTGAGATCAGACGTGATC---TTTAATGTGCGAAACCGGTGACATTTG	669
QY	133	SerSerAsnGluTyrIleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsn	152
DB	670	GGGTCAAGTCTTGGTGTGGAG-----	693
QY	153	ThrLeuLeuLysSerTyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAla	172
DB	694	---CTTACTGTGGCTCTTCATTACATTTCAATACTCCACAGACAAGATTCTTTGGGAT	750
QY	173	GlnGluGlyIleSerTyrTyrGluLysValLeuAlaLysTyrLysAspLeuGluSer	192
DB	751	GTGTGTCATCAGTCTTATCTCTCATAGATCTTACTGGGAGAGAGAAAGATGCCTACA	810
QY	193	IleLysLys	195
DB	811	ATCAGGCAA	819

#### RESULT 36

AAA61156  
ID AAA61156 standard; DNA; 2079 BP.

XX  
AC AAA61156;

XX  
DT 12-OCT-2000 (first entry)

XX  
DE SEN virus genome fragment SEQ ID NO: 103.

XX  
KW SEN virus; SENV; gastrointestinal tract disorder; inflammatory disease; proliferative disorder; hepatopathy; hepatitis; viral infection; vaccination; gene therapy; ds.

XX  
OS Hepatitis virus.

XX  
PN WO200028039-A2.

XX  
PD 18-MAY-2000.

XX  
PF 09-NOV-1999; 99WO-EP08566.

XX  
PR 10-NOV-1998; 98IT-MI02437.

XX  
PR 30-APR-1999; 99IT-MI00923.

XX  
PR 14-MAY-1999; 99EP-0830298.

XX  
PR 16-JUL-1999; 99EP-0113932.

XX  
PA (DIAS-) DIASORIN SRL.

XX  
PI Primi D, Fiordalisi G, Mantero GL, Mattioli S, Sottini A; Bonelli F, Vaglini L, Olivero P, Dal Corso A, Bonelli M;

XX  
DR WPI; 2000-376551/32.

XX  
PT Nucleic acids representing the genome of the SEN virus (SENV) and encoded proteins, useful for treatment of hepatopathies, inflammatory diseases and proliferative disorders such as cancer.

XX  
PS Example 16; Page 76-77; 392pp; English.

CC  
CC The present invention is concerned with the sequence of the genome of the SEN virus (SENV), and the proteins encoded by it. SENV is thought to be the cause of hepatopathies which are not linked to the presence of the hepatitis A, B and E viruses in man. The genome and proteins of this virus can be used in gene therapy and vaccination against the virus, which also causes disorders of the gastrointestinal tract, including Crohn's disease and lupus erythematosus, inflammatory diseases, and proliferative disorders such as cancer.

XX  
SQ Sequence 2079 BP; 656 A; 575 C; 432 G; 416 T; 0 other;

Alignment Scores:

Pred. No.: 8,95e-14 Length: 2079  
 Score: 277.50 Matches: 78  
 Percent Similarity: 52.00% Conservative: 13  
 Best Local Similarity: 44.57% Mismatches: 46  
 Query Match: 12.20% Indels: 39  
 DB: 21 Gaps: 6

US-10-057-531A-2 (1-431) x AAA61156 (1-2079)  
 Qy 1 MethHisHisHisHisHisHisSerSerGlyLeuValProArgGlySerGlyMetLysGlu 20  
 Db 1 ATGCACCATCATCATCATCTTCTGCTGTGGTCCACCGGGTCTGGTATGAAGAA 60  
 Qy 21 ThrAlaAlaLysPheGluArgGlnHisMetAspSerProAspLeuGlyThrAspAsp 40  
 Db 61 ACCGCTGCTGTAATTCGAACGCCAGCACATGGACAGCCAGATCTGGGTACCGAGCAG 120  
 Qy 41 AspAspLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMetAlaIleSer 60  
 Db 121 GACGACAAAGGCCATGGCTGATATCGGATCCATGAAC----- 156  
 Qy 61 ValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyAspValIleTyLeuLys 80  
 Db 157 -----TATGCCATGCACCTGCCGAGACAGACCCGCCAGAGAA 195  
 Qy 81 ProLeuAlaGlyValTyArgSerLeuLysGlnIleGluLysAsnIlePheThrPhe 100  
 Db 196 CCTTTCGGGGC-----GACATGTCACCAATAACTTTC 228  
 Qy 101 Asn---LeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyPheLeu 119  
 Db 229 AGCCTGCTGGTACTATACGACACGACAGACACCTTAACAGGTGGACCTTCCCAAC 288  
 Qy 120 AspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyIleIle 139  
 Db 289 GACCAGCTA-----GACCTGGTGGGTACAAACAC----- 318  
 Qy 140 GluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLysSerTyTyLys 159  
 Db 319 ---ACCAGGTTAAATTTTACAGAAGC-----AAAGACACTGA-CTTATAGTTACCTT 368  
 Qy 160 TyrlleLysGluSerValGluAsnAspIleLysPheAlaGlnGlu 174  
 Db 369 TAACATAAAGCCTCCCATGAATGAACGAGACTACTCTCACCAGAA 413

RESULT 37  
 ID AAA09770 standard; DNA; 7375 BP.  
 AC AAA09770;  
 XX  
 XX 23-JUN-2000 (first entry)  
 DT  
 DE Secreted protein acidic and rich in cysteine (SPARC) nucleotide sequence.  
 DE SPARC; secreted protein acidic and rich in cysteine; mouse; nerve cell;  
 KW drug; research reagent; mouse; ds.  
 KW Mus musculus.  
 OS  
 XX JP3012931-B1.  
 PN 28-FEB-2000.  
 PD 26-FEB-1999; 99JP-0049826.  
 XX 26-FEB-1999; 99JP-0049826.  
 PR 26-FEB-1999; 99JP-0049826.  
 XX (AGEN ) KOGYO GIJUTSUINCHO.  
 PA WPI; 2000-259560/23.  
 DR P-PSDB; AAY85150.  
 XX

PT New secretion protein acidic and rich in cysteine (SPARC) for drugs and  
 XX research reagents consists of specific amino acid sequence -  
 PS Claim 3; Page 8-15; 17pp; Japanese.  
 XX This sequence represents a nucleotide sequence encoding a mouse protein,  
 CC the protein has the physiological activity of SPARC (secreted protein  
 CC acidic and rich in cysteine). The protein is used for the preparation of  
 CC drugs, research reagents and pharmaceuticals. The SPARC protein is  
 CC prepared efficiently and inexpensively in large quantities. The protein  
 CC can be used in pharmaceuticals as it does not lose its physiological  
 CC activity.  
 XX  
 SQ Sequence 7375 BP; 1760 A; 1955 C; 1945 G; 1715 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 6,72e-13 Length: 7375  
 Score: 275.00 Matches: 50  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 98.04% Mismatches: 0  
 Query Match: 12.09% Indels: 0  
 DB: 21 Gaps: 0

US-10-057-531A-2 (1-431) x AAA09770 (1-7375)  
 Qy 1 MethHisHisHisHisHisHisSerSerGlyLeuValProArgGlySerGlyMetLysGlu 20  
 Db 5554 ATGCACCATCATCATCATCTTCTGCTGTGGTCCACCGGGTCTGGTATGAAGAA 5613  
 Qy 21 ThrAlaAlaLysPheGluArgGlnHisMetAspSerProAspLeuGlyThrAspAsp 40  
 Db 5614 ACCGCTGCTGTAATTCGAACGCCAGCACATGGACAGCCAGATCTGGGTACCGAGCAG 5673  
 Qy 41 AspAspLysAlaMetAlaAspIleGlySerIle 51  
 Db 5674 GACGACAAAGGCCATGGCTGATATCGGATCCATG 5706

RESULT 38  
 ID AAZ98759 standard; cDNA; 7375 BP.  
 XX AAZ98759;  
 AC AAZ98759;  
 XX 20-JUN-2000 (first entry)  
 DT  
 DE Murine SPARC nucleotide sequence.  
 DE Secreted protein acidic and rich in cysteine; SPARC; mouse; nerve cell;  
 KW neurocyte cell adhesion; cell protrusion retraction; cell shrinkage;  
 KW nervous system disease; epilepsy; arteriosclerosis; wound healing; ss.  
 OS Mus sp.  
 XX JP3012930-B1.  
 PN 28-FEB-2000.  
 PD 26-FEB-1999; 99JP-0049708.  
 XX 26-FEB-1999; 99JP-0049708.  
 PR (AGEN ) KOGYO GIJUTSUINCHO.  
 PA WPI; 2000-306484/27.  
 DR P-PSDB; AAY79067.  
 XX Drug composition for suppressing neurocyte cell adhesion, generating  
 PT cell migration and promoting shrinkage retraction of the cell contains  
 PT specific amino acid sequence  
 XX Example 2; Page 10-12; 21pp; Japanese.  
 XX This sequence represents a nucleotide sequence encoding a protein  
 CC

CC referred to as SPARC (secreted protein acidic and rich in cysteine). The protein is used as the active ingredient in a drug composition for suppressing neurocyte cell adhesion, generating cell migration and CC promoting a nerve protrusion shrinkage reaction. The composition can be used in wound healing and also to treat diseases of the nervous system, CC arteriosclerosis, and epilepsy.

XX Sequence 7375 BP; 1760 A; 1955 C; 1945 G; 1715 T; 0 other;

Alignment Scores:  
Pred. No.: 6,72e-13 Length: 7375  
Score: 275.00 Matches: 50  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 98.04% Mismatches: 0  
Query Match: 12.09% Indels: 0  
DB: 21 Gaps: 0

US-10-057-531A-2 (1-431) x AA298759 (1-7375)

QY 1 MetHisHisHisHisSerSerGlyLeuValProArgGlySerGlyMetLysGlu 20  
|||||  
Db 5554 ATGCACCATCATCATCATCTTCTGTGTCGTCACGCGGTTCTGTATGAAAGAA 5613  
QY 21 ThrAlaAlaLysPheGluArgGlnHisMetAspSerProAspLeuGlyThrAspAsp 40  
|||||  
Db 5614 ACCGCTGCTGCTAAATTCGAACGCCACGACATGACAGCCACAGATCTGGGTACCGACGAC 5673  
QY 41 AspAspLysAlaMetAlaAspIleGlySerIle 51  
|||||  
Db 5674 GAGCACAAAGCCATGCTGATATCGATCCATGATG 5706

RESULT 39

ID AAA94064  
ID AAA94064 standard; DNA; 241 BP.

XX AC AAA94064;

XX 30-JAN-2001 (first entry)

DE H. contortus clone 65e vector pET30a cloning junction DNA #2.

XX Nematode; parasite; helminth; sheep; goat; stomach; vaccine; ds.

XX Synthetic.

XX Key Location/Qualifiers

XX CDS 14..241

XX FT /\*tag- a

XX FT /product= "cloning junction sequence"

XX FT /partial

XX WO200056763-A1.

XX 28-SEP-2000.

XX 16-MAR-2000; 2000WO-AU00210.

XX 18-MAR-1999; 99AU-0009297.

XX (NOVS ) NOVARTIS AG.

XX (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.

XX Savin KW, Cook VR, Chen Y, Sexton JL, Apos E, Wilson LR;

XX Griffiths TM, Newton SE;

XX WPI: 2000-594573/56.

XX P-PSDB; AAB26110.

XX New Haemonchus contortus polypeptide for inducing a protective effect  
PT against a helminth by controlling helminth infection, growth, viability  
PT and/or egg fecundity and for ameliorating the symptoms of helminth  
PT infection -

XX

PS Disclosure; Fig 8; 94pp; English.

XX The present sequence comprises the sequence found at the cloning  
CC junction in the pET30a vector of nematode Haemonchus contortus  
CC clone 65e. This clone encodes a novel protein. This organism is a  
CC parasite found in the stomach of its host (sheep and goats in  
CC particular). The protein can be used in a vaccine against other  
CC helminths, as well as Haemonchus contortus, including trematodes,  
CC cestodes, nematodes and acanthocephala. These are all capable of causing  
CC severe illness in their hosts, which include sheep, pigs, goats, cattle,  
CC horses, donkeys, dogs, cats, guinea pigs and cage-birds, along with  
CC humans. Antibodies to the protein can be used to diagnose infection.  
CC Note: This sequence is stated as being the same as that shown in SEQ ID  
CC NO: 5 of the specification (see AAA94054). However, this sequence is  
CC shorter than the one shown here.

XX Sequence 241 BP; 69 A; 64 C; 59 G; 49 T; 0 other;

Alignment Scores:

Pred. No.: 1.29e-14 Length: 241  
Score: 274.00 Matches: 50  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 12.04% Indels: 0  
DB: 21 Gaps: 0

US-10-057-531A-2 (1-431) x AAA94064 (1-241)

QY 1 MetHisHisHisHisSerSerGlyLeuValProArgGlySerGlyMetLysGlu 20  
|||||  
Db 14 ATGCACCATCATCATCATCTTCTGTGTCGTCACGCGGTTCTGTATGAAAGAA 73

QY 21 ThrAlaAlaLysPheGluArgGlnHisMetAspSerProAspLeuGlyThrAspAsp 40  
|||||  
Db 74 ACCGCTGCTGCTAAATTCGAACGCCACGACATGACAGCCACAGATCTGGGTACCGACGAC 133

QY 41 AspAspLysAlaMetAlaAspIleGlySer 50

|||||

Db 134 GAGCACAAAGCCATGCTGATATCGGATCC 163

RESULT 40

AAAD18937/C

ID AAD18937 standard; DNA; 428 BP.

XX AC AAD18937;

XX 18-DEC-2001 (first entry)

XX Clone 6 DNA related to the invention.

XX Recombinant protein; dietary supplement; therapy; hypertension;

XX hypotensive; ds.

XX Unidentified.

XX WO200168674-A2.

XX 20-SEP-2001.

XX 13-MAR-2001; 2001WO-US07957.

XX 13-MAR-2000; 2000US-188990P.

XX (MONS ) MONSANTO TECHNOLOGY LLC.

XX Wang Q, Zhonghong G, Baggot BO, Hadfield K, Zhao J, Edwards J;

XX WPI: 2001-607456/69.

XX Producing polynucleotides having repeats, by annealing first  
PT polynucleotide pool having tandem repeats of sequences encoding  
PT recombinant protein with its complement pool, extending 3' ends and  
PT denaturing products -

XX Example 4; Fig 7; 77pp; English.

PS  
XX  
XX  
CC The invention relates to the field of recombinant DNA technology  
CC and, more particularly, to methods for the production of recombinant  
CC polynucleotide sequences and recombinant proteins containing repeating  
CC sequences. The invention provides methods for the production of  
CC polynucleotides and/or proteins containing tandem repeats of short  
CC sequences in which complementary polynucleotides anneal and act as  
CC primers allowing extension of their 3' ends to produce long sequences  
CC of tandem repeats. The method is useful for producing polynucleotide  
CC containing repeating units and for producing recombinant protein  
CC containing tandem units. The proteins produced can be used as dietary  
CC supplements in the treatment and control of hypertension. The present  
CC sequence is a polynucleotide produced by the method of the present  
CC invention from Clone 6.

XX  
SQ Sequence 428 BP; 85 A; 88 C; 119 G; 127 T; 9 other;

#### Alignment Scores:

Pred. No.:	2.59e-14	Length:	428
Score:	274.00	Matches:	50
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	12.04%	Indels:	0
DB:	22	Gaps:	0

US-10-057-531A-2 (1-431) x AAD18937 (1-428)

Qy	1	MethHisHisHisHisHisSerSerGlyLeuValProArgGlySerGlyMetLysGlu	20
Db	372	ATGCACCATCATCATCATCTTCTCTGGTCTGGTCCACCGGTTCTGGTATGAAGAA	313
Qy	21	ThrAlaAlaAlaLysPheGluArgGlnHisMetAspSerProAspLeuGlyThrAspAsp	40
Db	312	ACCGCTGCTGCTAAATTCGAACGCCAGCACATGGAGACGCCAGATCTGGGTACCGACGAC	253
Qy	41	AspAspLysAlaMetAlaAspIleGlySer	50
Db	252	GACGACAAGGCCATGGCTGATATCGGATCC	223

Search completed: May 19, 2003, 15:10:41  
Job time : 261 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 19, 2003, 15:04:10 ; Search time 47 Seconds  
(without alignments)  
2812.291 Million cell updates/sec

Title: US-10-057-531A-2

Perfect score: 2275

Sequence: 1 MHHHHHHSSGLVPRGSKGK.....TCECTKPDSPYLFDFGIFCSS 431

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_plus\_p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO\_spool/US10057531/runat\_12052003\_091032\_21701/app\_query.fasta\_1.583  
-DB=Issued\_Patents\_NA -OPMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -LOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=40  
-MODE=LOCAL -OUTFMT=pc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10057531 -CGN\_1\_1\_32\_runat\_12052003\_091032\_21701 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1935.5	85.1	5181	1	US-08-257-073-10 Sequence 10, Appl
2	1931	84.9	1219	4	US-08-195-705-1 Sequence 1, Appl
3	278.5	12.2	2472	4	US-09-626-589-6 Sequence 6, Appl
4	275	12.1	7375	4	US-09-513-442-1 Sequence 1, Appl
5	274	12.0	501	4	US-09-741-243C-3 Sequence 3, Appl
6	274	12.0	1653	4	US-09-741-243C-1 Sequence 1, Appl
7	259.5	11.4	1227	2	US-08-683-007A-1 Sequence 1, Appl
8	259.5	11.4	6619	2	US-08-683-007A-3 Sequence 3, Appl
9	242	10.6	587	4	US-09-166-966E-1 Sequence 1, Appl
10	173.5	7.6	1146	4	US-09-277-716-21 Sequence 21, Appl
11	173.5	7.6	1146	4	US-09-609-161B-21 Sequence 21, Appl
12	152.5	6.7	867	4	US-09-273-839A-7 Sequence 7, Appl

13	151	6.6	2136	3	US-08-946-475-8 Sequence 8, Appl
14	151	6.6	2136	4	US-09-340-479-8 Sequence 8, Appl
15	148.5	6.5	440	4	US-09-470-191-45 Sequence 45, Appl
16	139.5	6.1	3033	4	US-09-134-001C-2341 Sequence 2341, Ap
17	133	5.8	30549	4	US-09-134-001C-322 Sequence 322, App
18	131	5.8	3763	1	US-07-792-865D-1 Sequence 1, Appl
19	128	5.6	3883	2	US-08-468-036-33 Sequence 33, Appl
20	128	5.6	3883	2	US-08-376-843-33 Sequence 33, Appl
21	128	5.6	3884	4	US-09-541-782-3 Sequence 3, Appl
22	128	5.6	3884	4	US-09-723-820-3 Sequence 3, Appl
23	127.5	5.6	2836	4	US-09-157-257-3 Sequence 3, Appl
24	126.5	5.6	2637	4	US-09-134-001C-1541 Sequence 1541, Ap
25	126.5	5.6	6773	4	US-09-166-350-27 Sequence 27, Appl
26	126	5.5	3466	1	US-08-468-036-38 Sequence 38, Appl
27	126	5.5	3466	2	US-08-376-843-38 Sequence 113, App
28	126	5.5	3628	3	US-08-480-640A-113 Sequence 113, App
29	126	5.5	3628	3	US-08-295-802-113 Sequence 113, App
30	126	5.5	3628	4	US-08-488-237A-113 Sequence 113, App
31	126	5.5	3628	4	US-08-375-992A-113 Sequence 113, App
32	126	5.5	3942	3	US-08-480-640A-189 Sequence 189, App
33	126	5.5	3942	4	US-08-686-968C-189 Sequence 189, App
34	126	5.5	3942	4	US-08-488-237A-189 Sequence 189, App
35	126	5.5	3942	4	US-08-375-992A-189 Sequence 189, App
36	126	5.5	5785	3	US-08-480-640A-221 Sequence 221, App
37	126	5.5	5785	4	US-08-686-968C-221 Sequence 221, App
38	126	5.5	5785	4	US-08-488-237A-221 Sequence 221, App
39	125.5	5.5	7573	1	US-08-287-959-2 Sequence 2, Appl
40	125	5.5	2107	4	US-08-235-836C-73 Sequence 73, Appl
41	123	5.4	8133	1	US-09-134-001C-931 Sequence 931, Appl
42	122.5	5.4	8133	1	US-08-480-604A-5 Sequence 5, Appl
43	122.5	5.4	8133	2	US-08-405-496A-5 Sequence 5, Appl
44	122.5	5.4	8133	4	US-08-915-136-5 Sequence 5, Appl
45	122.5	5.4	8133	4	US-08-957-310-5 Sequence 5, Appl

ALIGNMENTS

RESULT 1  
US-08-257-073-10  
; Sequence 10, Application US/08257073  
; Patent No. 5765597  
; GENERAL INFORMATION:  
; APPLICANT: Paoletti, Enzo  
; APPLICANT: de Taisne, Charles  
; APPLICANT: Tine, John A.  
; TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Curtis, Morris & Safford, P.C.  
; STREET: 530 Fifth Avenue, 25th Floor  
; CITY: New York  
; STATE: NEW YORK  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/257,073  
; FILING DATE: 09-JUN-1994  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; PRIOR APPLICATION NUMBER: US 08/075,783  
; FILING DATE: 11-JUN-1993  
; PRIOR APPLICATION DATA:  
; PRIOR APPLICATION NUMBER: US 07/852,305  
; FILING DATE: 18-MAR-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/672,183  
; FILING DATE: 20-MAR-1991  
; ATTORNEY/AGENT INFORMATION:

NAME: Frommer, William S.  
 REGISTRATION NUMBER: 25,506  
 REFERENCE/DOCKET NUMBER: 454310-2570  
 TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 840-3333  
 TELEFAX: (212) 840-0712

TELEX: 425066 CURTMS

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 5181 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-257-073-10

Alignment Scores:

Pred. No.: 5,3e-201 Length: 5181  
 Score: 1935.50 Matches: 372  
 Percent Similarity: 96.45% Conservative: 8  
 Best Local Similarity: 94.42% Mismatches: 9  
 Query Match: 85.08% Indels: 5  
 DB: 1 Gaps: 1

US-10-057-531A-2 (1-431) x US-08-257-073-10 (1-5181)

Qy 38 ThrAspAspAspLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMet 57  
 Db 3952 TCCGAAGATAATGACGAATATTAGATCAAGTAGTAACTGGAGAA----- 3996  
 Qy 58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77  
 Db 3997 GCAATATCTGTCACATGGGAATATCTCTCAGGATTTTGAATAATGAATATGATGTTATA 4056  
 Qy 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97  
 Db 4057 TATTTAAACCTTTAGCTGGAGTATATAGAAGCTTAAACAAACAAATGAAACAACTT 4116  
 Qy 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117  
 Db 4117 TTTACATTTAATTTAAATTTTGAAGGATATCTTAAATTCAGCTCTTAAGAACGAAATAT 4176  
 Qy 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerAsnGluTyr 137  
 Db 4177 TTCTTAGATGATTAGAATCTGATTTAATGCAATTTAAACATATATCTCAATGAATAC 4236  
 Qy 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157  
 Db 4237 ATTATTGAAGATTCAATTTAAATTTAATTCAGAACAAACAAACACACATTTTAAAAAGT 4296  
 Qy 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177  
 Db 4297 TACAAATATATAAAGAATCAGTAGAATAATGATATTAATTTGCACAGGAGGTATTAAGT 4356  
 Qy 178 TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValIle 197  
 Db 4357 TATTATGAAGGTTTATAGCAATATATAGGATGATTAGATCAATTAACAAAGTTATC 4416  
 Qy 198 LysGluGluLysGluLysPheProSerSerProProThrThrProProSerProAlaLys 217  
 Db 4417 AAAGAAGAAAGGAGAGAGTTCCCATCATCATCACCACCAACACACCTCCGTCACCAGCAAA 4476  
 Qy 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237  
 Db 4477 ACAGACGAACAAAGAGAGAAAGTAAAGTCTCTCCATTTTAAACAAACATTTGAGACCTTA 4536  
 Qy 238 TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn 257  
 Db 4537 TACATAACTTAGTATATAAATGACGATTAATTAATTAACITTAAGCGCAAGATTAAAC 4596  
 Qy 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277  
 Db 4597 GATTGTAAATGTTGAAAGAGATGAAGCACATGTTAAAAATACTAAACCTTACTGATTTAAAA 4656

Qy 278 AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys 297  
 Db 4657 GCAATTGATGACAAATAGATCTTTTAAACCAATACGACTTCGAAGCAATTTAAAAA 4716  
 Qy 298 LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal 317  
 Db 4717 TTGATAATGATGATGACGAAAAAGATATGCTTGGCAAAATTTACTTAGTACAGGATTAGTT 4776  
 Qy 318 GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeu 337  
 Db 4777 CAAATTTTCTTAATAAATAATATCAAAATTAATTAAGGAAAAATTTCCAAGATATGTTA 4836  
 Qy 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357  
 Db 4837 AACATTTCAACACCAATCGTAAAAACAATGTCAGAAAAATTTCTGGATGTTTCAGA 4896  
 Qy 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377  
 Db 4897 CATTTAGATGAAGAGAGAAATGTAATGTTTATTAATTAACAAACAAAGAGGTGATAAA 4956  
 Qy 378 CysValGluAsnProAsnProThrCysAsnGluAsnGlyGlyCysAspAlaAspAla 397  
 Db 4957 TGTGTTGAAAAATCCAAATCTTACTTGTAAACGAAAAATTAATGTTGGATGTCAGATGCC 5016  
 Qy 398 ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417  
 Db 5017 AAATGTACCGAAGAAGATTCAGGTAGCAACGAAAGAAAAATCACATGTGAATGTACTAAA 5076  
 Qy 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSer 431  
 Db 5077 CCGATTCCTTATCCACTTTTCGATGGTAAATTTCTGCAGTTCC 5118  
 RESULT 2  
 US-08-195-705-1  
 : Sequence 1, Application US/08195705  
 : Patent No. 6420523  
 : GENERAL INFORMATION:  
 : APPLICANT: Chang, Sandra  
 : APPLICANT: Hul, George  
 : APPLICANT: Barr, Philip  
 : APPLICANT: Gibson, Helen  
 : TITLE OF INVENTION: RACULOVIRUS PRODUCED PLASMIDIUM  
 : TITLE OF INVENTION: FALCIPARUM VACCINE  
 : NUMBER OF SEQUENCES: 5  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Davis Hoxie Faithfull Hapgood  
 : STREET: 45 Rockefeller Pl.  
 : CITY: New York  
 : STATE: N.Y.  
 : COUNTRY: USA  
 : ZIP: 10111  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: PatentIn Release #1.0, Version #1.25  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/195,705  
 : FILING DATE: 14-FEB-1994  
 : CLASSIFICATION: 435  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Jacobs, Seth H  
 : REGISTRATION NUMBER: 32140  
 : REFERENCE/DOCKET NUMBER: 11880A3  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: 212-757-2200  
 : TELEFAX: 212-586-1461  
 : INFORMATION FOR SEQ ID NO: 1:  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 1219 base pairs  
 : TYPE: nucleic acid  
 : STRANDEDNESS: double  
 : TOPOLOGY: linear



```
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Plasmodium falciparum
: STRAIN: falciparum uganda palo alto
: IMMEDIATE SOURCE:
: CLONE: flg5LFUP42AcNPV
: FEATURE:
: NAME/KEY: exon
: LOCATION: 13..1194
: US-08-195-705-1

Alignment Scores:
Pred. No.: 1219
Score: 1.78e-201
Percent Similarity: 98.67%
Best Local Similarity: 98.40%
Query Match: 84.88%
DB: 4

US-10-057-531A-2 (1-431) x US-08-195-705-1 (1-1219)

QY 56 ThrMetAlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAsp 75
Db 70 ACCGCGCGGATATCTGTCACAATGGATTAATATCTCTCAGGATTTGAAATGAATATGAT 129
QY 76 ValIleTyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLys 95
Db 130 GTTATATATTTAAACCTTTAGCTGGAGTATATAGAGCTTTAAACCAATTTGAAAAA 189
QY 96 AsnIlePheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArg 115
Db 190 AACATTTTTCATTTAAATTTGAACGATATCTTAAATTCACGCTCTTAAAGAACGA 249
QY 116 LysTyrPheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsn 135
Db 250 AAATATTTCTTAGATGATTTAGATCTGATTTAATGCAATTTAAACATATATCTCTCAAT 309
QY 136 GluTyrIleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeu 155
Db 310 GAATACATTTTGAAGATTCATTTAAATTTAATTTGCAACACCAACCAACACACATTTTA 369
QY 156 LysSerTyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGly 175
Db 370 AAAAGTTACAAATATATAAAGAAATCAGTAGAAATGATATAATTTGACAGGAAGGT 429
QY 176 IleSerTyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLys 195
Db 430 ATAGTTTATTTAGAAAGGTTTTCGCAATATAAGGATGATTTAGAAATCAATTTAAACAA 489
QY 196 ValIleLysGluGluLysGluLysPheProSerSerProProThrThrProProSerPro 215
Db 490 GTTATCAAGAACAAAGGAGAGTCCCATCATCATCACCAACCAACACCTCCCTCACCA 549
QY 216 AlaLysThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGlu 235
Db 550 GCAAAACACAGCAACCAAGAAAGGAAAGTAAAGTTCCTTCCATTTTAAACAAATTTAG 609
QY 236 ThrLeuTyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLys 255
Db 610 ACCTTATACAAATTAACCTAGTTAATAAATTTGACGATTTACTTAATTAACCTTAAGGCA 669
QY 256 IleAsnAspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAsp 275
Db 670 ATTAACGATTTGATTTGTTGAAAGATGAACACATGTTAAATATACTAACTTACTGAT 729
QY 276 LeuLysAlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIle 295
Db 730 TTAAGCAATTTGATGACAAATATAGATCTTTTAAAAAACCATTAACGACTTCGAGCAATT 789
QY 296 LysLysLeuIleAsnAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGly 315
Db 406 ACCGCTGCTGCTAAATTCGAACGCCAGCACATGGACAGCCAGATCTGGGTACCGACGAC 465

Db 790 AAAAAATTGATAATGATGATACGAAAAAAGATATGCTGTGCAAAATTTACTTAGTACAGGA 849
QY 316 LeuValGlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGluAsp 335
Db 850 TTAGTTCAAAATTTTCTTAATACAAATATATCAAAATTAATTTGAAGGAAATTTCCAAGAT 909
QY 336 MetLeuAsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCys 355
Db 910 ATGTTAAACATTTTCACAACACCAATGCTAAAAAACAATGTCTCCAGAAAAATTTCTGGATGT 969
QY 356 PheArgHisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGly 375
Db 970 TTCAGACATTTTAGATGAAGAGAGAAATGTAATGTTTATTAATTTCAACAAGAAGGT 1029
QY 376 AspLysCysValGlnAsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAla 395
Db 1030 GATAAATGCTGTTGAAATCCAAATCCTACTTTGTAACCAAAATAATGTTGGATGTGATGCA 1089
QY 396 AspAlaThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCys 415
Db 1090 GATGCCAAATGTACCGAAGAGATTCAGGTAGCAACGAGAAAGAAATCACATGTGAATGT 1149
QY 416 ThrLysProAspSerTyrProLeuPheAspGlyIlePheCysSer 430
Db 1150 ACTAAACCTGATTTCTTATCCACTTTTCGATGGTATTTTCTGCAGT 1194

RESULT 3
US-09-626-589-6
: Sequence 6, Application US/09626589
: Patent No. 6326164
: GENERAL INFORMATION:
: APPLICANT: Rice, John
: APPLICANT: Kloti, Andreas
: APPLICANT: Crawford, John
: APPLICANT: Lanning, Beth
: APPLICANT: Stewart, Sandy
: TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
: IDENTIFICATION OF MODULATORS OF DEOXYXYLULOSE 5-PHOSPHATE
: FILE REFERENCE: 2037 US
: CURRENT APPLICATION NUMBER: US/09/626, 589
: CURRENT FILING DATE: 2000-07-27
: NUMBER OF SEQ ID NOS: 6
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 6
: LENGTH: 2472
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Nucleotides 1-495 encode the thioredoxin sequence
: found in the vector pET32 supplied by No. 6326164agen.
: OTHER INFORMATION: Nucleotides 496-2472 represent the tDXPS cDNA
: OTHER INFORMATION: sequence from Arabidopsis thaliana.
US-09-626-589-6

Alignment Scores:
Pred. No.: 1.67e-20
Score: 278.50
Percent Similarity: 49.75%
Best Local Similarity: 37.44%
Query Match: 12.24%
DB: 4

US-10-057-531A-2 (1-431) x US-09-626-589-6 (1-2472)

QY 1 MetHisHisHisHisHisSerSerGlyLeuValProArgGlySerGlyMetLysGlu 20
Db 346 ATGCACCATCATCATCATCTTCTGCTGTGTCACCGCGGTCTGCTGATGAAAGAA 405
QY 21 ThrAlaAlaLysPheGluArgGlnHisMetAspSerProAspLeuGlyThrAspAsp 40
Db 406 ACCGCTGCTGCTAAATTCGAACGCCAGCACATGGACAGCCAGATCTGGGTACCGACGAC 465
```

```
Qy 41 AspAspLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMetAlaIleSer 60
Db 466 GACGACAAAGGCCATGCTGATACGGATCCGCTTCA-----501
Qy 61 ValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyraAspValIleTyrlLeuLys 80
Db 501 -----501
Qy 81 ProLeuAla-----GlyValTyraArgSerLeuLysLysGlnIleGluLysAsnIlePhe 98
Db 502 ---CTTGACAGAGCGGTGAATATTATTC---AACAGACCACCACTCCATTACTTGAC 555
Qy 99 ThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyrrPhe 118
Db 556 ACTATTAACCTACCCCAATCCAC---ATGAAATCTTCTGTCAAGGAACCTGAAACAACATT 612
Qy 119 LeuAspValLeuGluSerAspLeuMetGlnPheLys-----HisIle 132
Db 613 TCTGATGAGCTGAGATCAGAGTGATC---TTTANTGTCTGMAAACCGGTGGACATTGG 669
Qy 133 SerSerAsnGluTyrlleIleGluAspPheLysLeuLeuAsnSerGluGlnLysAsn 152
Db 670 GGGTCAAGCTTGTGTGTGGAG-----693
Qy 153 ThrLeuLeuLysSerTyrlLysTyrlLysGluSerValGluAsnAspIleLysPheAla 172
Db 694 ---CTTACTGGCTTCATTACATTTTCAATCTCCACAGACAAGATTCTTTGGGAT 750
Qy 173 GlnGluGlyIleSerTyrrTyrrGluLysValLeuAlaLysTyrrLysAspLeuGluSer 192
Db 751 GTTGTCATCAGCTTATCCCTCATAAGATTCTTACTGGGAGAAGAGAAAGATGCTTACA 810
Qy 193 IleLysLys 195
Db 811 ATGAGGCAA 819

RESULT 4
US-09-513-442-1
; Sequence 1, Application US/09513442
; Patent No. 6387664
; GENERAL INFORMATION:
; APPLICANT: Ikemoto, Mitsushi
; TITLE OF INVENTION: SPARC Fusion Protein and a Process for Producing the
; TITLE OF INVENTION: Same
; FILE REFERENCE: HIRAKI-04218
; CURRENT APPLICATION NUMBER: US/09/513.442
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 7375
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (5209)..(6609)
US-09-513-442-1

Alignment Scores:
Pred. No.: 2.15e-19 Length: 7375
Score: 275.00 Matches: 50
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 98.04% Mismatches: 0
Query Match: 12.09% Indels: 0
DB: 4 Gaps: 0

US-10-057-531A-2 (12431) x US-09-513-442-1 (1-7375)

Qy 1 MethHisHisHisHisHisSerSerGlyLeuValProArgGlySerGlyMetLysGlu 20
Db 5554 ATGCACCATCATCATCATCTTCTGTGTCGCCACGCGTCTGTGATGAAGAA 5613
Qy 21 ThrAlaAlaAlaLysPheGluArgGlnHisMetAspSerProAspLeuGlyThrAspAsp 40
```

```
Db 5614 ACCGCTGCTGCTAAATTCGACGCCAGCACATGGACGCCAGATCTGGTATCCGACGAC 5673
Qy 41 AspAspLysAlaMetAlaAspIleGlySerIle 51
Db 5674 GACGACAAAGGCCATGCTGATATCGGATCCCATG 5706

RESULT 5
US-09-741-243C-3
; Sequence 3, Application US/09741243C
; Patent No. 6399352
; GENERAL INFORMATION:
; APPLICANT: Crawford Jr., John Milton
; APPLICANT: Rice, John
; APPLICANT: Sevala, Veeresh
; APPLICANT: Stewart, Sandy
; TITLE OF INVENTION: NOVEL PLANT PORPHOBILINOGEN AND FUSION
; TITLE OF INVENTION: PROTEIN THEREOF
; FILE REFERENCE: 20220S
; CURRENT APPLICATION NUMBER: US/09/741.243C
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/171,785
; PRIOR FILING DATE: 1999-12-22
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 501
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: DNA encoding thioredoxin functional fragment
US-09-741-243C-3

Alignment Scores:
Pred. No.: 4.44e-21 Length: 501
Score: 274.00 Matches: 50
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 12.04% Indels: 0
DB: 4 Gaps: 0

US-10-057-531A-2 (12431) x US-09-741-243C-3 (1-501)

Qy 1 MethHisHisHisHisHisSerSerGlyLeuValProArgGlySerGlyMetLysGlu 20
Db 346 ATGCACCATCATCATCATCTTCTGTGTCGCCACGCGTCTGTGATGAAGAA 405
Qy 21 ThrAlaAlaAlaLysPheGluArgGlnHisMetAspSerProAspLeuGlyThrAspAsp 40
Db 406 ACCGCTGCTGCTAAATTCGACGCCAGCACATGGACGCCAGATCTGGTATCCGACGAC 465
Qy 41 AspAspLysAlaMetAlaAspIleGlySer 50
Db 466 GACGACAAAGGCCATGCTGATATCGGATCC 495

RESULT 6
US-09-741-243C-1
; Sequence 1, Application US/09741243C
; Patent No. 6399352
; GENERAL INFORMATION:
; APPLICANT: Crawford Jr., John Milton
; APPLICANT: Rice, John
; APPLICANT: Sevala, Veeresh
; APPLICANT: Stewart, Sandy
; TITLE OF INVENTION: NOVEL PLANT PORPHOBILINOGEN AND FUSION
; TITLE OF INVENTION: PROTEIN THEREOF
; FILE REFERENCE: 20220S
; CURRENT APPLICATION NUMBER: US/09/741.243C
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/171,785
; PRIOR FILING DATE: 1999-12-22
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
```





```
; Sequence 21, Application US/09277716A
; Patent No. 6232107
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROLUME, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
; CURRENT APPLICATION NUMBER: US/09/277,716A
; EARLIER FILING DATE: 1999-03-26
; EARLIER FILING DATE: 1998-10-01
; EARLIER FILING DATE: 1998-10-01
; EARLIER FILING DATE: 1998-06-15
; EARLIER FILING DATE: 1998-06-15
; EARLIER FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 1146
; TYPE: DNA
; ORGANISM: Artificial Sequence: fusion construct
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1146)
; OTHER INFORMATION: DNA encoding a Cellulose Binding Domain (CBD)-Gaussia luciferase
US-09-277-716-21

Alignment Scores:
Pred. No.:      Length:      1146
Score:          173.50      Matches:      35
Percent Similarity: 92.31%      Conservative: 1
Best Local Similarity: 89.74%      Mismatches: 0
Query Match:      7.63%      Indels:      3
DB:               4      Gaps:      1

US-10-057-531A-2 (1-431) x US-09-277-716-21 (1-1146)
; Sequence 21, Application US/09609161B
; Patent No. 6436682
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROLUME, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE LUC
; TITLE OF INVENTION: AND FLUORESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HI
; FILE REFERENCE: 24729-121B
; CURRENT APPLICATION NUMBER: US/09/609,161B
; PRIOR FILING DATE: 1999-03-26
; PRIOR FILING DATE: 1998-10-01
; PRIOR FILING DATE: 1998-10-01
; PRIOR FILING DATE: 1998-06-15
; PRIOR FILING DATE: 1998-06-15
; PRIOR FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 1146
; TYPE: DNA
```

```
; ORGANISM: Artificial Sequence: fusion construct
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1146)
; OTHER INFORMATION: DNA encoding a Cellulose Binding Domain (CBD)-Gaussia lucifera
US-09-609-161B-21

Alignment Scores:
Pred. No.:      Length:      1146
Score:          173.50      Matches:      35
Percent Similarity: 92.31%      Conservative: 1
Best Local Similarity: 89.74%      Mismatches: 0
Query Match:      7.63%      Indels:      3
DB:               4      Gaps:      1

US-10-057-531A-2 (1-431) x US-09-609-161B-21 (1-1146)
QY 8 SerSerGlyLeuValProArgGlySer-----GlyMetLysGluThrAlaAla 24
Db 475 TCCGCGGGTCTGGTGCCACGCGGTACTGCAATTGGTATGAAGAAACCGCTGCTGCT 534
QY 25 LysPheGluArgGlnHisMetAspSerProAspLeuGlyThrAspAspAspLys 43
Db 535 AAATTGCAAGCCGACACATGGACGCCAGATCTGGTACCGATGACGACACAAG 591

RESULT 12
US-09-273-839A-7
; Sequence 7, Application US/09273839A
; Patent No. 6329156
; GENERAL INFORMATION:
; APPLICANT: Cirino, Nick M
; APPLICANT: Jackson, Paul J
; APPLICANT: Lehnert, Bruce E
; TITLE OF INVENTION: Disruption of Anthrax Toxin Binding to Cell Surface
; FILE REFERENCE: S-89,662
; CURRENT APPLICATION NUMBER: US/09/273,839A
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 867
; TYPE: DNA
; ORGANISM: Bacillus anthracis
US-09-273-839A-7

Alignment Scores:
Pred. No.:      Length:      867
Score:          152.50      Matches:      74
Percent Similarity: 40.88%      Conservative: 65
Best Local Similarity: 21.76%      Mismatches: 96
Query Match:      6.70%      Indels:      105
DB:               4      Gaps:      19

US-10-057-531A-2 (1-431) x US-09-273-839A-7 (1-867)
QY 2 HisHisHisHisHisSerSerGlyLeuValProArgGlySerGlyMetLysGlu--- 20
Db 13 CATCATCATCATCATCACAGCGCCCTGGTGCCCGCGGCGCATATGCTCGAGGCG 72
QY 21 ---ThrAlaAlaLysPheGluArgGlnHisMetAspSerProAspLeuGlyThrAsp 39
Db 73 GCGCATGCCGCCGAAACAACTGCACGTATCATTTTAAATGGAAGAAATTAATCTG--- 129
QY 40 AspAspAspLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMetAla 59
Db 130 ---GTAGAAGCGGATAGCGCGGTAAATCTCTAGTATGATCCATTAGAACGACTAAACCG 186
QY 60 SerValThrMetAspAsnIleLeuSer-----GlyPheGlu----- 71
Db 187 GATATGACATTAAAGAACGCCCTTAAATAGCATTTGGATTTAAACGACCAATGGAAC 246
```

```
Qy 72 -----AsnGluTyrAspValIleTyrLeuLysProLeuAla 83
Db 247 TTACAATATCAAGGAAAGACATAACCAATTTGATTTTAAATTCGATCAACAACATCT 306
Qy 84 GlyValTyrArgSerLeuLysGlnIleGluLys -----AsnIlePheThr 99
Db 307 -----CAAAATATCAAGAAATCAGTTACCGGAATTAACGCACTAATATATAT 417
Qy 100 Phe -----AsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117
Db 358 GTATTAGATAAAATCAAAATTAATGCAAAATGAATATTTTAATAGAGATAAACGTTT 417
Qy 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137
Db 418 CATTATGAT-----AGAAATAAATAGCA-----GTTGGCGCGATGAGTCA 459
Qy 138 IleIleGluAspSerPheLys---LeuLeuAsnSerGluGlnLysAsnThrLeuLys 156
Db 460 GTAGTTAAGGAGGCTCATAGAGAAGTAATTAATTCGTCAACAGAGGATATTGTTA--- 516
Qy 157 SerTyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyLe 176
Db 517 -----AATATTGATAAGGATATAAGA----- 537
Qy 177 SerTyrTyrGluLysValLeuAlaLysTyr-----LysAspAspLeuGluSerIle 193
Db 538 -----AAAATATTATAGGTTATATCTGAGAAATGAAGATCTGAAGGCTT 585
Qy 194 LysLysValIleLysGluGluLysPheProSerSerProThrProThrProPro 213
Db 586 AAAGAAGTTATAATGACAGATATGATGTTGAATTTCT----- 627
Qy 214 SerProAlaLysThrAspGluGlnLysLysGluSerLys---PheLeuProPheLeuThr 232
Db 628 -----AGTTTACGCGCAAGATGCAAAACATTATATAGATTTT----- 663
Qy 233 AsnIleGluThrLeuTyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeu 252
Db 664 -----AAAAATAATAT-----GATAAATTACGTTATATATAAGTAATCCC 705
Qy 253 LysAlaLysIleAsnAspCysAsnValIleLysAspGluAlaHisValLysIleThrLys 272
Db 706 AATTATAAGGTAATGATATGCTGTACTATAAGAAACACTATATT----- 753
Qy 273 LeuSerAspLeuLysAlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPhe 292
Db 754 -----AATCCTAGTGAG----- 765
Qy 293 GluAlaIleLysLysLeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeu 312
Db 766 -----AATGGGATACTAGTACCACCGGATCAAGAAATTTTA 804

RESULT 13
US-08-946-475-8
: Sequence 8, Application US/08946475
: Patent No. 6013505
: GENERAL INFORMATION:
: APPLICANT: Gwynn, Michael
: APPLICANT: Kallender, Howard
: APPLICANT: Palmer, Leslie
: TITLE OF INVENTION: TOPOISOMERASE I
: NUMBER OF SEQUENCES: 13
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SmithKline Beecham Corporation
: STREET: 709 Swedeland Road
: CITY: King of Prussia
: STATE: PA
: COUNTRY: USA
: ZIP: 19406-0939
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
```

```
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/946,475
: FILING DATE: 08-OCT-1997
: CLASSIFICATION: 435
```

```
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 60/027,973
: FILING DATE: 08-OCT-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Gimmi, Edward R
: REGISTRATION NUMBER: 38,891
: REFERENCE/DOCKET NUMBER: P50560
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 610-270-4478
: TELEFAX: 610-270-5090
: TELEX:
```

```
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2136 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: Genomic DNA
: US-08-946-475-8
```

```
Alignment Scores:
Pred. No.: 1,12e-06 Length: 2136
Score: 151.00 Matches: 68
Percent Similarity: 36.03% Conservativeness: 39
Best Local Similarity: 22.90% Mismatches: 80
Query Match: 6.64% Indels: 110
DB: 3 Gaps: 13
```

US-10-057-531A-2 (1-431) x US-08-946-475-8 (1-2136)

```
Qy 2 HisHisHisHisHisHisSerSerGlyLeuValProArgGlySerGlyMet----- 18
Db 13 CATCATCATCATCATCAGCAGCGCGCTGGTCCGCGCGCGCATATGACCTTGGCA 72
Qy 19 -----LysGluThrAlaAlaLysPheGluArgGlnHisMetAsp 32
Db 73 GATAATTTAGTCATTTGTAATCCCTGCAAAACCAAAACCATTTGAAGATATTAGGT 132
Qy 33 SerProAspLeuGlyThrAspAspAspLysAlaMetAlaAspIleGlySerIleGlu 52
Db 133 AAG-----AAATATAAAGTTTAGCTTCAATGGGACACGTCAGA 171
Qy 53 Gly-----ArgGlyThrMetAlaIleSerValThrMetAspAsnIleLeuSerGlyPhe 70
Db 172 GACTTACCAGAGAGTCAATGGGTGTCGAC---ACTGAAGATAAT-----TAC 216
Qy 71 GluAsnGluTyrAspValIleTyrLeuLysProLeuAlaGlyValTyrArgSerLeuLys 90
Db 217 GAACCAAAATAT-----ATAACAATACGCGGAAAGGCTCTGTGTGTAAGAAGATTGAAA 270
Qy 91 LysGlnIleGlu-----LysAsnIlePheThrPheAsn----- 101
Db 271 AAACATGCAAAAAAAGCGGAAAGAGCTCTTCTCGCAAGTGACCCCGCGGTGAAGGTGAA 330
Qy 102 -----LeuAsnLeuAsnAspIleLeuAsnSerArgLeu 112
Db 331 GCATTTGCTTGGCATTTTATCAAAATTTTAGAGCTTGAAGATTTCAAGAAATTCGCGTT 390
Qy 113 LysLysArgLysTyrPheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIle 132
Db 391 GTTTTCAACGAAATACTAAAGAGCGCTGTTAAAGAAAGTTTAAAGAAATCTAGAGAATT 450
Qy 133 SerSerAsnGluTyrIleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsn 152
Db 451 GAAATGAAC-----TTAGTCGATGCACAAACACGCGT 483
Qy 153 ThrLeuLeuLysSer-----TyrLysTyrIleLys 162
```

Db 484 CGAATATTAGATAGATTGTTGGCTATACATCTCCGCCAGTTCTATGGAGAAAAAGTGAA 543  
Qy 163 GluSerVal-----Glu 166  
Db 544 AAAGTTGTCAGCGGTCGAGTTCAATCTGTTGCATCTCGTTTAGTCATTGACCGTGAA 603  
Qy 167 AsnAspIle----- 169  
Db 604 AATGAATTCGAAACTTTAAACCAGAGAATATTGGACTATTGAAGGAGAATTTAGATAC 663  
Qy 170 -----LysPheAlaGlnGluGlyIleSerTyr----- 178  
Db 664 AAAAAATCAAAATCAATGCTAAATTCCTCATTAATAAAATAAAACCTTTTAAATTAATA 723  
Qy 179 -----TyrGluLysValLeuAlaLysTyrLysAspLeuGluSerIleLys 194  
Db 724 ACGAAAAAGATGTTGAGAAAAATTACAGCTGCATAGATGGAGATCAATTCGAAATTACA 783  
Qy 195 LysValIleLysGluLysGluLysPheProSerSerProThrThr 211  
Db 784 AACGTGACTAAAAAGAAAAACGGCTAATCCAGCAAAACCCATTACAACT 834

## RESULT 14

US-09-340-479-8

; Sequence 8, Application US/09340479

; Patent No. 6274139

; GENERAL INFORMATION:

; APPLICANT: Gwynn, Michael

; APPLICANT: Kallender, Howard

; APPLICANT: Palmer, Leslie

; TITLE OF INVENTION: TOPOISOMERASE I

; NUMBER OF SEQUENCES: 13

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SmithKline Beecham Corporation

; STREET: 709 Swedeland Road

; CITY: King of Prussia

; STATE: PA

; COUNTRY: USA

; ZIP: 19406-0939

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/340,479

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/946,475

; FILING DATE: 08-OCT-1997

; APPLICATION NUMBER: 60/027,973

; FILING DATE: 08-OCT-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Gimmli, Edward R

; REGISTRATION NUMBER: 38,891

; REFERENCE/DOCKET NUMBER: P50560

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 610-270-4478

; TELEFAX: 610-270-5090

; TELEX:

; INFORMATION FOR SEQ ID NO: 8:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2136 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: Genomic DNA

US-09-340-479-8

Alignment Scores:

Pred. No.: 1.12e-06

Score: 2136

Length:

Matches: 68

Percent Similarity: 36.03% Conservative: 39  
Best Local Similarity: 22.90% Mismatches: 80  
Query Match: 6.64% Indels: 110  
DB: 4 Gaps: 13

US-10-057-531a-2 (1-431) x US-09-340-479-8 (1-2136)

Qy 2 HisHisHisHisHisSerSerGlyLeuValProArgGlySerGlyMet----- 18  
Db 13 CATCATCATCATCATCATCAGCAGCGCCCTGGTCCCGCGGCGGACCATATGACCTTGGCA 72  
Qy 19 -----LysGluThrAlaAlaLysPheGluGluArgGlnHisMetAsp 32  
Db 73 GATAATTTAGTCATCTGTAATCGCTGCAAAAGCAAAACCATTTGAAAGATTTAGGT 132  
Qy 33 SerProAspLeuGlyThrAspAspAspLysAlaMetAlaAspIleGlySerIleGlu 52  
Db 133 AG-----AAATAAAGTTTAGCTTCAATGGACACAGTCAGA 171  
Qy 53 Gly-----ArgGlyThrMetAlaIleSerValThrMetAspAsnIleLeuSerGlyPhe 70  
Db 172 GACTTACCAAGAGTCAATGGGTGCAC---ACTGAAGATAAT-----TAC 216  
Qy 71 GluAsnGluTyrAspValIleTyrLeuLysProLeuAlaGlyValTyrArgSerLeuLys 90  
Db 217 GAACCAAAATAT-----ATAACATACCGGAAAGGTCCTGTTGTAAGAAGAAATTGAA 270  
Qy 91 LysGlnIleGlu-----LysAsnIlePheThrPheAsn----- 101  
Db 271 AACATGCAAAAAAGCGAAAAACGCTTTTCGCAAGATGACCCCGACCGTGAAGGTGAA 330  
Qy 102 -----LeuAsnLeuAsnAspIleLeuAsnSerArgLeu 112  
Db 331 GCAATTTGCTTGGCATTATCAAAATTTTAGAGCTTTGAAGATCTTAAGAAATTCGCGTT 390  
Qy 113 LysLysArgLysTyrPheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIle 132  
Db 391 GTTTTCAACGAAATAACTAAAGACGCTGTTAAAGAAAGTTTAAAAATCTCAGAGAAATT 450  
Qy 133 SerSerAsnGluTyrIleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsn 152  
Db 451 GAATGAAC-----TTAGTCGATGCACCAACGCGCGT 483  
Qy 153 ThrLeuLeuLysSer-----TyrLysTyrIleLys 162  
Db 484 CGAATATTAGATAGATTGTTGGCTATACATCTGCCAGTTCTATGGAAAAAGTGAA 543  
Qy 163 GluSerVal-----Glu 166  
Db 544 AAAGTTTTCAGCGGGTCGAGTTCAATCTGTTGCATCTCGTTTAGTCATTGACCGTGAA 603  
Qy 167 AsnAspIle----- 169  
Db 604 AATGAATTCGAAACTTTAAACCAGAGAATATTGGACTATTGAAGGAGAATTTAGATAC 663  
Qy 170 -----LysPheAlaGlnGluGlyIleSerTyr----- 178  
Db 664 AAAAAATCAAAATTCATGCTAAATTCCTTCAATATAAAAAATAAACCTTTTAAATTAATA 723  
Qy 179 -----TyrGluLysValLeuAlaLysTyrLysAspLeuGluSerIleLys 194  
Db 724 ACGAAAAAGATGTTGAGAAAAATTACAGCTGCATAGATGGAGATCAATTCGAAATTACA 783  
Qy 195 LysValIleLysGluLysGluLysPheProSerSerProThrThr 211  
Db 784 AACGTGACTAAAAAGAAAAACGGCTAATCCAGCAAAACCCATTACAACT 834

## RESULT 15

US-09-470-191-45/c

; Sequence 45, Application US/09470191

; Patent No. 6465633

; GENERAL INFORMATION:

; APPLICANT: Skeiky, Yasir

```

: APPLICANT: Corixa Corporation
: TITLE OF INVENTION: Compositions and Methods of Their Use in
: TITLE OF INVENTION: the Treatment, Prevention and Diagnosis of Tuberculosis
: FILE REFERENCE: 014058-008910US
: CURRENT APPLICATION NUMBER: US/09/470,191
: CURRENT FILING DATE: 1999-12-23
: PRIOR APPLICATION NUMBER: US 60/113,952
: PRIOR FILING DATE: 1998-12-24

```

Alignment Scores:	
Pred. No.:	1.86e-07
Score:	148.50
Length:	440
Matches:	43
Percent Similarity:	61.33%
Best Local Similarity:	57.33%
Conservative:	13
Mismatches:	3
Indels:	17
Query Match:	3
DB:	4
US-10-057-531A-2 (1-431) x US-09-470-191-45 (1-440)	



```

; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 322
; LENGTH: 30549
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-322

Alignment Scores:
Pred. No.: 0.0062 Length: 30549
Score: 133.00 Matches: 73
Percent Similarity: 37.96% Conservative: 50
Best Local Similarity: 22.53% Mismatches: 115
Query Match: 5.85% Indels: 87
DB: 4 Gaps: 14

US-10-057-531A-2 (1-431) x US-09-134-001C-322 (1-30549)
QY 12 valproargGlySerGlyMetLysGluThrAlaAlaLysPheGluArgGlnHisMet 31
||||| ||||| : : : : :
Db 12859 GTACCTAGTACAGATGCTATGACTCAGCAATCTGTTGCAAAATATATCAAAACTACAA 12918
: : : : :
QY 32 -----AspSerProAspLeuGly 37
: : : : :
Db 12919 ATAGCTAAAAACCAATTAACACAAATTAATACGTTTACCGAACAAATCCAGATGTTAAT 12978
: : : : :
QY 38 ThrAspAspAspLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMet 57
: : : : :
Db 12979 GCAATCAAAACGAATAAGCAGAAGCGAA----- 13008
: : : : :
QY 58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77
: : : : :
Db 13009 -----CGAATCAGTAACGATTTAAACACAGCTAAGATAACTTACAAAGTT--- 13053
: : : : :
QY 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97
: : : : :
Db 13054 GATACCTCAACCTTTAGAAAAAATAAAGACAACTTCAAGATGAAATGTATCAAGGTACT 13113
: : : : :
QY 98 PheThrPheAsnLeu-----AsnLeuAsnAspIleLeu----- 108
: : : : :
Db 13114 AACACAGATGCAATGACTCAAGATTCAGTGGATAATTACAATGATAGCTTAAGTGCAGCA 13173
: : : : :
QY 109 -----AsnSerArgLeuLysLysArgLysTyrPheLeuAspVal 121
: : : : :
Db 13174 ATTATAGAAAAGCAAGTAATAATTAATTAAGCGTAATCCGACAGTAGAACAAGTT 13233
: : : : :
QY 122 LeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyrIleIleGluAsp 141
: : : : :
Db 13234 AAAGAGAGC-----GTTGCTAATGCACAAACAAGTCATACAGAT 13272
: : : : :
QY 142 SerPheLysLeuLeuAsnSer-----GluGlnLysAsn 152
: : : : :
Db 13273 TTACAAAATGCTCAACTTCACTTGTTCCAGACAAAACCTCAACTTCAAGAGCTAAATAAT 13332
: : : : :
QY 153 ThrLeuLysSerTyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAla 172
: : : : :
Db 13333 AGATTAGAAAACAGT-----ATTACCAACAAACAGATAGTAC---GGCATGACT 13380
: : : : :
QY 173 GlnGluGlyIleSerTyrTyrGluLysValLeuAlaLysTyrLysAspLeuGluSer 192
: : : : :
Db 13381 CAAGATTCGCTTACATATTATATGATAAATGACAAACCTAGACAAACCTTGAAAAA 13440
: : : : :
QY 193 IleLysLysValIleLysGluLysGluLysPheProSerProThrThrPro 212

```

```

Db 13441 ATATCTAAAGTTTAA-----GGTGGCTCAACCTACTGTAGCT 13476
|||||
QY 213 ProSerProAlaLysThrAspGluGlnLysLysSerLysPheLeuProPheLeuThr 232
|||||
Db 13477 GAAATTAGACAAAATACAGATGAAGCAAAATGCACATAAACAAGCATTA----- 13524
: : : : :
QY 233 AsnIleGluThrLeuTyrAsnAsnLeuValAsnLysIleAspAspTyrLeu-----Ile 250
: : : : :
Db 13525 -----GACACTGCACGCTTCTCAACTTACATTAATAGAGAGCCATATATCAATCATATT 13578
: : : : :
QY 251 AsnLeuLysAlaLysIleAsnAspCysAsnValGluLysAspGluAlaHisValLysIle 270
: : : : :
Db 13579 AATAATGAAAGTCATTTAAAT-----AACGCGCAAAAGATAATTTTAAAGCTCAAGTT 13632
: : : : :
QY 271 ThrLysLeu-----SerAspLeuLysAlaIleAspAspLysIleAspLeuPheLys 287
: : : : :
Db 13633 AACTCAGCACCTAATCATATACTTTAGAAACGATTAAATAAGCTGTATAC-TTTAAA 13691
: : : : :
QY 288 AsnProTyrAsp 291
: : : : :
Db 13692 TCAATCTATGAC 13703

RESULT 18
US-07-792-865D-1
; Sequence 1, Application US/07792865D
; Patent No. 5646247
; GENERAL INFORMATION:
; APPLICANT: John W. Barnwell, Mary W. Gallinski,
; APPLICANT: Samuel P. Wertheimer
; TITLE OF INVENTION: MEROZOITE ANTIGENS LOCALIZED AT
; TITLE OF INVENTION: THE APICAL END OF THE PARASITE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby P.C.
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage
; COMPUTER: IBM or IBM-compatible
; OPERATING SYSTEM: PC/MS-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/792,865D
; FILING DATE: 19911004
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US90/01849
; FILING DATE: April 3, 1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson, Arthur
; REGISTRATION NUMBER: 34,354
; REFERENCE/DOCKET NUMBER: 5986/14692-US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 527-7700
; TELEFAX: (212) 753-6237
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3763 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; DESCRIPTION:
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM: P.vivax
; STRAIN:

```



QY 427 IlePheCys 429  
 Db 3705 ATAAATGC 3713

## RESULT 19

US-08-468-036-33  
 ; Sequence 33, Application US/08468036  
 ; Patent No. 5728806  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DeMaggio, Anthony J.  
 ; APPLICANT: Hoeksra, Merl F.  
 ; TITLE OF INVENTION: Materials and Methods Relating to Proteins that  
 ; INTERACT WITH CASEIN KINASE I  
 ; NUMBER OF SEQUENCES: 48  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
 ; STREET: 6300 Sears Tower, 233 South Wacker Drive  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: United States of America  
 ; ZIP: 60606-6402  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/468,036  
 ; FILING DATE: 06-JUN-1995  
 ; CLASSIFICATION: 530  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/184,605  
 ; FILING DATE: 21-JAN-1994  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: No. 5728806and, Greta E.  
 ; REGISTRATION NUMBER: 35,302  
 ; REFERENCE/DOCKET NUMBER: 27866/31784  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 312/474-6300  
 ; TELEFAX: 312/474-0448  
 ; TELEX: 25-3856  
 ; INFORMATION FOR SEQ ID NO: 33:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3883 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; US-08-468-036-33

Alignment Scores:  
 Pred. No.: 0.000916 Length: 3883  
 Score: 128.00 Matches: 89  
 Percent Similarity: 39.07% Conservative: 54  
 Best Local Similarity: 24.32% Mismatches: 109  
 Query Match: 5.63% Indels: 114  
 DB: 1 Gaps: 21

US-10-057-531a-2 (1-431) x US-08-468-036-33 (1-3883)

QY 86 TyrArgSerLeuLysLysGlnIleGluLysAsnIlePheThrPheAsnLeuAsn 105  
 Db 2011 TATGCTTCGAAGGCTAAACATTAAGAACACCGCAACTGGGTTTCATTATAATGAAG 2070  
 QY 106 AspIleLeuAsnSerArgLeuLysLysArgLysTyrPheLeuAspValLeuGluSerAsp 125  
 Db 2071 GATATTG-----GTTAAATAATACTATGGAATTAGCAAGATTAAATCCGAT 2121  
 QY 126 LeuMetGlnPheLysHisIleSerSerAsnGlu-----TyrIleIleGluAspSerPhe 143  
 Db 2122 TTACTCTCTCAAAAG-----TCCAAAGAGGAATATATATGAGCCAGATCACTAC 2172  
 QY 144 LysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSerTyrLysTyrIleLysGlu 163

Db 2173 AAAAATTTGAACAGTGAT-----TTAGAAAGTTATAAAATGAAGTTCAA 2217  
 QY 164 SerValGluAsnAspIleLysPheAlaGlnGluClyIleSerTyrTyrGluLysValLeu 183  
 Db 2218 GAATGTAAAAGAGAAATTGAA-----AGTTTGACATCGAAAAAATGCA 2259  
 QY 184 AlaLysTyrLysAspLeuGluSerIleLysLysValIleLysGluLysGluLys 203  
 Db 2260 TTGCTAGTAAAGATAAATTGAAGTCA--AAAGAACTATTCAATCTCAAAATGCGAA 2316  
 QY 204 PheProSerSerProProThrThrProProSerProAlaLysThrSerGluGlnLysLys 223  
 Db 2317 ATGAATCATTTGAAAACTACCATAGATCATTTAAGGCGACAACTAGATAAACAACATAAA 2376  
 QY 224 GluSerLysPheLeuProPheLeuThrAsnIleGlu---ThrLeuTyrAsnAsnLeuVal 242  
 Db 2377 -----ACTGAAATTGAAATATCCGATTTTTAATAACAACTA 2412  
 QY 243 AsnLysIleAspAspTyr-----LeuIleAsnLeuLysAlaLysIleAsnAsp 258  
 Db 2413 CAGAAGTTGACTGAGGTAATGCAAAATGCCCTACATGATTACAAAAAAGAGAACTTGAC 2472  
 QY 259 CysAsnValGluLysAspGluAlaHisValLysIleThrLys---LeuSerAspLeuLys 277  
 Db 2473 CTTAAT---CAAAAGTTTGAATGTCAT-----ATTACTAAAGAAATTAATAAATTCGAA 2523  
 QY 278 Ala----- 278  
 Db 2524 TCTACACTGTTTTCACAAATTAAACACTATGCAACAGAGAAAGTATTCTTCAAGAGACTAAT 2583  
 QY 279 IleAspAspLysIleAspLeuPheLysAsnPro----- 289  
 Db 2584 ATCCAACCAAAATCTTGATATGATCAAAAATGAAGTACTGACTCTTATGAGAACCATGCAA 2643  
 QY 290 -----TyrAspPheGluAlaIleLysLysLeuIleAsnAspAspThr 303  
 Db 2644 GAAAAAGCTGAACATAATGTACAAA---GACTGTGTGAAGAAATTTTAAACGAATCTCCT 2700  
 QY 304 LysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuValGlnAsnPheProAsnThr 323  
 Db 2701 AAA-----TTCTTCAATGTT 2715  
 QY 324 IleIleSerLys-----LeuIleGluGlyLysPheGln----- 334  
 Db 2716 GTTATTGAGAAATCGACATAATAAGAGTAGATTTCACAAAATTTTATAAAATATAGCC 2775  
 QY 335 ---AspMetLeuAsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSer 353  
 Db 2776 GAGAATCTTCTGTATATTAGCGAAGAAATAACAAACATGAACAGTAGTACTTAAAAAACCAT 2835  
 QY 354 GlyCysPhe-----ArgHisLeuAspGluArgGlu 363  
 Db 2836 -----TTTTTCAAGATAACCATCAAGAATTACTGAATGTCATGTGGATTCTACTTAT 2889  
 QY 364 GluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLysCysValGluAsnProAsn 383  
 Db 2890 GAA-----AATATTGACAAGAGAAACACGAGTTTGTGTGAGAACTTTTAAA 2934  
 QY 384 ProThrCysAsnGluAsn 389  
 Db 2935 AAGGTCTCTAAATGACCAC 2952

## RESULT 20

US-08-376-843-33  
 ; Sequence 33, Application US/08376843  
 ; Patent No. 5846764  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DeMaggio, Anthony J.  
 ; APPLICANT: Hoeksra, Merl F.  
 ; TITLE OF INVENTION: Materials and Methods Relating to Proteins  
 ; INTERACT WITH CASEIN KINASE I  
 ; NUMBER OF SEQUENCES: 53

;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
;; STREET: 6300 Sears Tower, 233 South Wacker Drive  
;; CITY: Chicago  
;; STATE: Illinois  
;; COUNTRY: United States of America  
;; ZIP: 60606-6402  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/376,843  
;; FILING DATE:  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/184,605  
;; FILING DATE: 21-JAN-1994  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: No. 5846764and, Greta E.  
;; REGISTRATION NUMBER: 35,302  
;; REFERENCE/DOCKET NUMBER: 27866/31784  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 312/474-6300  
;; TELEFAX: 312/474-0448  
;; TELEX: 25-3856  
;; INFORMATION FOR SEQ ID NO: 33:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 3883 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
US-08-376-843-33

## Alignment Scores:

Pred. No.: 0.000916 Length: 3883  
Score: 128.00 Matches: 89  
Percent Similarity: 39.07% Conservativity: 54  
Best Local Similarity: 24.32% Mismatches: 109  
Query Match: 5.63% Indels: 114  
DB: 2 Gaps: 21

US-10-057-531A-2 (1-431) x US-08-376-843-33 (1-3883)

Qy 86 TyrArgSerLeuLysLysGluGlnLeuLysAsnIlePheThrPheAsnLeuAsnLeuAsn 105  
Db 2011 TATGCTTCGAGGCTAAACACATTAAGACACAGCCGCAACTGGGTTCATTTATATGAG 2070  
Qy 106 AsnLeuAsnSerArgLeuLysLysArgLysTyrPheLeuAspValLeuGluSerAsp 125  
Db 2071 GATATTTTG-----GTTAAAAATATACTATGGAATTAGCAAGATTAATCCGAT 2121  
Qy 126 LeuMetGlnPheLysHisIleSerSerAsnGlu-----TyrIleIleGluAspSerPhe 143  
Db 2122 TTACTCTCTCAAG-----TCCAAAGAAAGAAATATATATGACCAAGATCACTAC 2172  
Qy 144 LysLeuLeuAsnSerGluGlnLysAsnThrLeuLysSerTyrLysTyrIleLysGlu 163  
Db 2173 AAAAAATTGACAGTGAT-----TTAGAAAGCTTATAAAATCAAGTTCAA 2217  
Qy 164 SerValGluAsnAspIleLysPheAlaGlnGluGlyIleSerTyrTyrGluLysValLeu 183  
Db 2218 GAATGTAAGAGAAATTGAA-----ACTTTGACATCCGAAAAATGCA 2259  
Qy 184 AlalysTyrLysAspAspLeuSerIleLysValIleLysGluGluLysGluLys 203  
Db 2260 TTGCTAGTAAAGATAAAATGAAGTCA---AAAGAACTATTCAATCTCAAAATGCGCAA 2316  
Qy 204 PheProSerSerProThrThrProProSerProAlaLysThrAspGluGlnLysLys 223  
Db 2317 ATAGAATCATTTGAAACTACCATAGATCATTTAAGGGCACAACTAGATAAACAGCATAA 2376

Qy 224 GluSerLysPheLeuProPheLeuThrAsnIleGlu-----ThrLeuTyrAsnAsnLeuVal 242  
Db 2377 -----ACTGAATTGAAATATCCGATTTTAAATAACAACAACTA 2412  
Qy 243 AsnLysIleAspAspTyr-----LeuIleAsnLeuLysAlaLysIleAsnAsp 258  
Db 2413 CAGAAGTTGACTAGGTAAATGCAATGGCCCTACATATTACAAAAAAGAGAACTTGAC 2472  
Qy 259 CysAsnValGluLysAspGluAlaHisValLysIleThrLys---LeuSerAspLeuLys 277  
Db 2473 CTTAAT---CAAAAGTTGAAATGCAT-----ATTACTAAGAAATTTAAAAAATTGAA 2523  
Qy 278 Ala----- 278  
Db 2524 TCTACACTGTTTTTACAATTAACACTATGCAACAGGAAAGATTAATCTTCAAGAGACTAAT 2583  
Qy 279 IleAspAspLysIleAspLeuPheLysAsnPro----- 289  
Db 2584 ATCCAACCAAAATCTTGATATGATCAAAATGAAGTACTGACTCTTATGAGAACCATGCAA 2643  
Qy 290 -----TyrAspPheGluAlaIleLysLysLeuIleAsnAspAspThr 303  
Db 2644 GAAAAGCTGAACATAATGTACAAA---GACTGTGTGGAAGAAATTTTAAACGAATCTCCT 2700  
Qy 304 LysLysAspMetLeuGlyLysLeuSerThrGlyLeuValGlnAsnPheProAsnThr 323  
Db 2701 AAA-----TTCTCAATGTT 2715  
Qy 324 IleIleSerLys-----LeuIleGluGlyLysPheGln----- 334  
Db 2716 GTTATTGAGAAAAATCGACATAATAAGAGTAGATTTCCTCAAAATTTTATAAAATATAGCC 2775  
Qy 335 ---AspMetLeuAsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSer 353  
Db 2776 GAGATCTTTCTGATATTAGCGAAGAAATAACACATGAAACAGTACTTTAAAAAACCAT 2835  
Qy 354 GlyCysPhe-----ArgHisLeuAspGluArgGlu 363  
Db 2836 -----TTTTTCAAGAAATAACCATCAAGAAATTACTGTAATCGTCATGTGATCTACTTAT 2889  
Qy 364 GluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLysCysValGluAsnProAsn 383  
Db 2890 GAA-----AATATTGAGAAAGAGAAACAACAGAGTTGTTGAGAACTTTAAA 2934  
Qy 384 ProThrCysAsnGluAsn 389  
Db 2935 AAGTCTCTAATGACCAC 2952

## RESULT 21

US-09-541-782-3  
; Sequence 3, Application US/09541782  
; Patent No. 6284480  
; GENERAL INFORMATION:  
; APPLICANT: Nislow, Corey  
; APPLICANT: Sakowicz, Roman  
; APPLICANT: Beraud, Christophe  
; TITLE OF INVENTION: Antifungal Assay  
; FILE REFERENCE: 1015  
; CURRENT APPLICATION NUMBER: US/09/541,782  
; CURRENT FILING DATE: 2000-04-03  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 3884  
; TYPE: DNA  
; ORGANISM: Saccharomyces cerevisiae  
US-09-541-782-3

Alignment Scores:  
Pred. No.: 0.000916 Length: 3884  
Score: 128.00 Matches: 89  
Percent Similarity: 39.07% Conservativity: 54

Best Local Similarity: 24.32% Mismatches: 109  
Query Match: 5.63% Indels: 114  
DB: 4 Gaps: 21

US-10-057-531A-2 (1-431) x US-09-541-782-3 (1-3884)

QY 86 TyrArgSerLeuLysGlnIleGluLysAsnIlePheThrPheAsnLeuAsnLeuAsn 105  
DB 2012 TATGCTCGAAGCTAAACATTAAAGCAAGCGCAACTGGGTTCATTATATGAAG 2071

QY 106 AspIleLeuAsnSerArgLeuLysArgLysTyrPheLeuAspValLeuGluSerAsp 125  
DB 2072 GATATTTG-----GTTAAATAATACTATGGAATTAGCAACAGATTAAATCCGAT 2122

QY 126 LeuMetGlnPheLysHisIleSerSerAsnGlu-----TyrIleIleGluAspSerPhe 143  
DB 2123 TTACTCTCTACAAG-----TCCAAAGAAGAAATATATAGCCAAAGATCACATAC 2173

QY 144 LysLeuLeuAsnSerGluGlnLysAsnThrLeuLysSerTyrLysTyrIleLysGlu 163  
DB 2174 AAAAATTTGAACACTGAT-----TTAGAAAGTTATATAAATCAAGTTCAA 2218

QY 164 SerValGluAsnAspIleLysPheAlaGlnGluLysTyrTyrGluLysValLeu 183  
DB 2219 GAATGTAAAGAGAAATTGAA-----AGTTTGACATCGAAAAATGCA 2260

QY 184 AlaLysTyrLysAspAspLeuGluSerIleLysLysValIleLysGluLysGluLys 203  
DB 2261 TTGCTAGTAAAGATTAATTTGAAGTCA---AAAGAAGTTTCAATCTCAAAATTCGCAA 2317

QY 204 PheProSerSerProThrThrProProSerProAlaLysThrAspGluLysLys 223  
DB 2318 ATAGAATCATTTGAAACTACCATAGATCATTTTAAAGGCACAACTAGTAACAGCATAAA 2377

QY 224 GluSerLysPheLeuProPheLeuThrAsnIleGlu---ThrLeuTyrAsnAsnLeuVal 242  
DB 2378 -----ACTGAAATTTGAATATCCGATTTTAAATAACAACTA 2413

QY 243 AsnLysIleAspAspTyr-----LeuIleAsnLeuLysAlaLysIleAsnAsp 258  
DB 2414 CAGAAGTTGACTGAGGTAATCAAAATGCCCTACATGATTACAAAAAAGAGAACTTGAC 2473

QY 259 CysAsnValGluLysAspGluAlaHisValLysIleThrLys---LeuSerAspLeuLys 277  
DB 2474 CTTAAT---CAAAAGTTTGAATGCAT-----ATTACTAAAGAAATTTAAAAAATTGAAA 2524

QY 278 Ala----- 278

DB 2525 TCTACACTGTTTTTACAATTAACACTATGCAACAGAGAAAGATTCTTCAAGAGACTAAT 2584

QY 279 IleAspAspLysIleAspLeuPheLysAsnPro----- 289

DB 2585 ATCCAACCAATCTTGATATATCATCAAAATGAAGTACTGACTCTTATGAGAACCATGCAA 2644

QY 290 -----TyrAspPheGluAlaIleLysLysLeuIleAsnAspAspThr 303  
DB 2645 GAAAAAGCTGAATAAGTACAAA---CAGTGTGTGAAGAAATTTTAAACGAATCTCTCT 2701

QY 304 LysLysAspMetLeuGlyLysLeuSerThrGlyLeuValGlnAsnPheProAsnThr 323  
DB 2702 AAA-----TTCTTCAATGTT 2716

QY 324 IleIleSerLys-----LeuIleGluGlyLysPheGln----- 334

DB 2717 GTTATGTGAGAAATTCGACATAATTAAGAGTAGATTTCCTCAAAATTTTATAAAATATAGCC 2776

QY 335 ---AspMetLeuAsnIleSerGlnHisLysValLysLysGlnCysProGluAsnSer 353  
DB 2777 GAGAAATCTTCTGATATTAGCGAAGAAATTAACACATCAACAGACTTCTTAAAAACCAT 2836

QY 354 GlyCysPhe-----ArgHisLeuAspGluArgGlu 363  
DB 2837 -----TTTTTCAAGAAATAACCATCAAGAAATTTACTGAATTCGTGTTGTTCTACTTAT 2890





Score: 126.50 Matches: 83  
Percent Similarity: 40.82% Conservative: 66  
Best Local Similarity: 22.74% Mismatches: 119  
Query Match: 5.56% Indels: 97  
DB: 4 Gaps: 22

US-10-057-531A-2 (1-431) x US-09-166-350-27 (1-6773)

Qy 63 MetAspAsnIleLeuSerGlyPheGluAsnGluTyArgValIleTyLeuLysProLeu 82  
|||||  
Db 1778 ATGGATAATTTCCATAAGAAATGTGAAGGCGAAGAGATTGATCTT----- 1825

Qy 83 AlaGlyValTyArgSerLeuLysGlnIleGluLysAsnIlePheThrPheAsnLeu 102  
|||||  
Db 1826 -----GAACTTGGGAAGAAAGTAGAGCAACAATC---CAGTCAACAAGT 1867

Qy 103 AsnLeu-----AsnAspIle-----LeuAsnSerArgLeuLysLysArg 115  
|||||  
Db 1868 GACTAGACAAAAGGTAAATGAATTAACAGGAGGACTAGAGGAGCTTTAAAGAAAG 1927

Qy 116 LysTyArgPheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsn 135  
|||||  
Db 1928 GATCAAAATGACCAAAACTAGAAAACCTTATGCTTCAATGAAAGTCTCTCTGAAGAC 1987

Qy 136 GluTyIleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeu 155  
|||||  
Db 1988 AAAGAAGTATTTGTCAGCTGAAGTCACTCTCT---TATGAGGAAACAAATAAATCACTCAGT 2044

Qy 156 LysSerTyIleLysLysValGlu-----AsnAspIle 169  
|||||  
Db 2045 TCAGAAAAAACAGTTGAGTAGGATTTGGAGGTTTTTTGCTCAAAAAGAGAGTGT 2104

Qy 170 LysPheAlaGlnGluGlyIleSerTyTyArgLysValLeuAlaLysTyIleAspAsp 189  
|||||  
Db 2105 ---ATCCTTAAGAACATATTAATCACTTACAAAGAAAGTTCAGTTAATGTTGAAGAG 2161

Qy 190 LeuGluSerIleLysLysValIleLysGlu----- 200  
|||||  
Db 2162 CAAGATAATTTAAATAAACTGCTGAAATAGCAAGTTCAGAGATTTATTTGTTAAACT 2221

Qy 201 -----LysGluLysPheProSerSerProThrProProSer 214  
|||||  
Db 2222 CAGTTGTATGTTTCTTAAGAAATG-----GGATCA 2254

Qy 215 ProAlaLysThrAspGluGlnLysLysLysSerLysPheLeuProPheLeuThrAsnIle 234  
|||||  
Db 2255 GAAGTTTTCAGAAAGACAGTGAAGAGAAAGAT----- 2284

Qy 235 GluThrLeuTyArgAsnLeuValAsnLysIleAspAspTyIleLeuAsnLeuLysAla 254  
|||||  
Db 2285 -----GTTGTAATGCTTACAGCAGCTCGTGAATCCCTG-----GCA 2323

Qy 255 LysIleAsnAsp-----CysAsnVal-----GluLysAspGluAlaHisValLysIle 270  
|||||  
Db 2324 AAATAAATGAGAAAATGCAACCTGCTTTTCAGCGTGATGAGAAAGTATTAGAGTTA 2383

Qy 271 ThrLysLeuSerAspLeuLysAlaIleAspLysIleAspLeuPheLysAsnProTy 290  
|||||  
Db 2384 GAAAA-----GAGATTAAAGTGCCTTCAAGAA-----GAGAGTGTAGTT 2422

Qy 291 AspPheGluAlaIleLysLysIleAsnAsp---AspThrLysLysAspMetLeuGly 309  
|||||  
Db 2423 CAGTGTGAGAACATTAGCTCTTTTATGAGACATATGAGCAAGAGAAAGTCTCTTAAGG 2482

Qy 310 LysLeuLeuSerThrGlyLeuValGlnAsnPheProAsnThrIleIleSerLysLeuIle 329  
|||||  
Db 2483 AAAGAGTTAGAGAA-----ATACAGTCAGAAAAAGAGGCCCTGACATCTGATCTTCTA 2536

Qy 330 GluGlyLys-----PheGlnAspMetLeu----- 337  
|||||  
Db 2537 GAAATGAAGAAATGCTAATGAAAAAACAGGCTTGAAATCAAGAACTTTTAAATCAAGTT 2596

Qy 338 ---AsnIleSerGlnHisGlnCysValLysLysGlnCys---ProGluAsnSerGlyCys 355  
|||||

Db 2597 GAAGAAGTATCTCAA---ACATGTAGCAAAAGTGAATTCCTAATGAAAAAGAAAAATGT 2653  
|||  
Qy 356 PheArgHisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyIleGlnGluGly 375  
|||  
Db 2654 TTT-----ATAAGGAAACATGAAACCTAAAGCCACTACTAGAACAAAAAGAAATTAGA 2707

Qy 376 AspLysCysValGlu 380  
|||  
Db 2708 GATAGGAGAGCAGAG 2722

RESULT 26  
US-08-468-036-38  
; Sequence 38, Application US/08468036  
; Patent No. 5728806  
; GENERAL INFORMATION:  
; APPLICANT: DeHaggio, Anthony J.  
; APPLICANT: Hoekstra, Merl F.  
; TITLE OF INVENTION: Materials and Methods Relating to Proteins that  
; TITLE OF INVENTION: Interact with Casein Kinase I  
; NUMBER OF SEQUENCES: 48  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/468,036  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/184,605  
; FILING DATE: 21-JAN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: No. 5728806and, Greta E.  
; REGISTRATION NUMBER: 35,302  
; REFERENCE/DOCKET NUMBER: 27866/31784  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/474-6300  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 38:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3466 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-08-468-036-38

Alignment Scores:  
Pred. No.: 0.00127 Length: 3466  
Score: 126.00 Matches: 73  
Percent Similarity: 41.94% Conservative: 70  
Best Local Similarity: 21.41% Mismatches: 126  
Query Match: 5.54% Indels: 72  
DB: 1 Gaps: 15

US-10-057-531A-2 (1-431) x US-08-468-036-38 (1-3466)

Qy 70 PheGluAsnGluTyArgValIleTyLeuLysProLeuAlaGlyValTyArgSerLeu 89  
|||  
Db 1235 TTCTAATGATGAGACATGAAATCAATTTAAGCAGCTAGCGCATCTAATAAAAAAGAG 1294

Qy 90 LysLysGlnIleGluLysAsnIlePheThrPheAsnLeuAsnAspIleLeuAsn 109  
|||||



Db 1295 CTGAACAAATGGAATAATGAA-----TACAAACAAAAATGAGAAATTTGAAATTT 1345  
QY 110 SerArgLeuLys-----LysArgLysTyrPheLeuAspValLeuGlu----- 123  
Db 1346 ATCAAGATTAAACAGTTTGAAGAGCGTCGCTTTACATAAATAGAAGAGGTA 1405  
QY 124 -----SerAspLeuMetGln 128  
Db 1406 AGAAATAAATACCATGAACCTTCCACTTTACAGGAATGTTGAACGATGTTGAACAA 1465  
QY 129 PheLysHisIleSerSerAsnGluTyrIleGlu---AspSerPheLysLeuLeuAsn 147  
Db 1466 AACCATATGCTTGAAGAGAGAGTGCCTTACAGATACCAATCGCAGTGAAGAGAGGAT 1525  
QY 148 SerGluGlnLysAsnThrLeuLeuLysSerTyrLysTyrIleLysGluSerValGluAsn 167  
Db 1526 ATAGAGCTGAATATAAACAATATCAAGAAATCGAAAGCATAAAGAGGAAATCGAAAT 1585  
QY 168 AspIleLysPheAlaGlnGluGlyIleSerTyrTyrGluLysValLeuAlaLysTyrLys 187  
Db 1586 ACATTAATA-----CCTGAGTTGGCAGAAAAAGAGAGCTTAAACAGAAAGCGT 1636  
QY 188 AspAspLeuGluSerIleLysValIleLysGluGluLysGluLysPheProSerSer 207  
Db 1637 AACCGGTATGAGCTATCAAGTTAAAGTTAAAGAAAGAGAGGAA----- 1684  
QY 208 ProProThrThrProProSerProAlaLysThrAspGluGlnLysLysGluSerLysPhe 227  
Db 1685 -----ACTACAGGCTGAGAGATGAGGTGCGCATTTAAACACAGAA--- 1723  
QY 228 LeuProPheLeuThrAsnIleGluThrLeuTyrAsnAsnLeuValAsnLysIleAspAsp 247  
Db 1724 -----ACTAATTTAGAACTTTC-----GAAAGATCAAGAACTTGAGGAA 1765  
QY 248 TyrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsn-----ValGluLysAspGlu 265  
Db 1766 TATATAAGACACTGAGCTGGTGTGAAGGAGTTGAATGAAATTCGTATTAAGAGGAA 1825  
QY 266 AlaHisValLysIleThrLysLeuSerAspLeuLysAlaIleAspAspLysIleAspLeu 285  
Db 1826 ACG---GTTAGCGCACATTGCTATATGAGTTTACAGAGTTAAGAGGAAATATACGAGTT 1882  
QY 286 Phe-----LysAsnProTyrAspPheGluAlaIleLysLysLysLeuLeuAsnAspAspThr 303  
Db 1883 TATTGTAGGATTCGTCCA-----GCTCTAAAAAATTTGGAAAAATTCGTACT 1930  
QY 304 LysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuValGlnAsnPheProAsnThr 323  
Db 1931 AGCCTTATTATGTTAATGATTTGATGACAATAGTGGTGTTCATCTATGGAAGTGACG 1990  
QY 324 IleIle-----SerLysLeuIleGluGlyLysPheGlnAspMetLeuAsnIleSer 340  
Db 1991 AAAATACAAAACACAGCGCAAGTGCATGATTCAAAATTTTCAATAAATATTGAT----- 2044  
QY 341 GlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArgHisLeuAsp 360  
Db 2045 -----CAACAGGATACAAATGTGATGTTTAAAGAAAGTGGT 2083  
QY 361 GluArgGluGluCysLys-----CysLeuLeuAsnTyrLysGlnGlu 374  
Db 2084 CAGTTAGTGCAAAGTTCAATTAGATGGATATATGTTTGTATCTTCGCATACGCNCAACA 2143  
QY 375 Gly 375  
Db 2144 GGA 2146

## RESULT 27

US-08-376-843-38

; Sequence 38, Application US/08376843

; Patent No. 5846764

; GENERAL INFORMATION:

; APPLICANT: DeMaggio, Anthony J.

; APPLICANT: Hoekstra, Meri F.

; TITLE OF INVENTION: Materials and Methods Relating to Proteins  
; TITLE OF INVENTION: that Interact with Casein Kinase I  
; NUMBER OF SEQUENCES: 53  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/376,843  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/184,605  
; FILING DATE: 21-JAN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: No. 5846764and, Greta E.  
; REGISTRATION NUMBER: 35,302  
; REFERENCE/DOCKET NUMBER: 27866/31784  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/474-6300  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 38:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3466 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-08-376-843-38

## Alignment Scores:

Score: 0.00127 Length: 3466  
Percent Similarity: 126.00 Matches: 73  
Best Local Similarity: 41.94% Conservativeness: 70  
Query Match: 21.41% Mismatches: 126  
Indels: 72  
DB: Gaps: 15

US-10-057-531A-2 (1-431) x US-08-376-843-38 (1-3466)

QY 70 PheGluAsnGluTyrAspValIleTyrLeuLysProLeuAlaGlyValTyrArgSerLeu 89  
Db 1235 TTCATGAATGAAGGACATGAATACATTAAGAGCAGCTAGCGGCATCTAATAAAAAAGAG 1294  
QY 90 LysLysGlnIleGluLysAsnIlePheThrPheAsnLeuAsnLeuAsnAspIleLeuAsn 109  
Db 1295 CTGAACAAATGGAATAATGAA-----TACAAACAAAAATGAGAAATTTGAAATTT 1345  
QY 110 SerArgLeuLys-----LysArgLysTyrPheLeuAspValLeuGlu----- 123  
Db 1346 ATGAGATTAAACAGTTTGAAGAGGAGCGTCGCTTTAGATAAATAAGAGAGGTA 1405  
QY 124 -----SerAspLeuMetGln 128  
Db 1406 AGAAATAAATCACCATGAACCTTCCACTTTACAGGAATGTTGAACGATGTTGAACAA 1465  
QY 129 PheLysHisIleSerSerAsnGluTyrIleIleGlu---AspSerPheLysLeuLeuAsn 147  
Db 1466 AACCATATGCTTGAAGAGAGAGTGCCTTACAGATACCAATCGCAGTGAAGAGAGGAT 1525  
QY 148 SerGluGlnLysAsnThrLeuLeuLysSerTyrLysTyrIleLysGluSerValGluAsn 167  
Db 1526 ATAGAGCTGAATATAAACAATATCGAAAGATCGAAAGAGGAAATCGAAAT 1585



```
QY 221 GlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrAsnAsn 240
    : : : : :
Db 2355 ATTATAAGATATATAAAATTATTCAGT-----AAAAATAGATAAAGCAT 2399
    : : : : :
QY 241 LeuValAsnLysIleAspSerTyrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsn 260
    : : : : :
Db 2400 GAAAATAATAAGGTGGAGGAGGTGTGTACATATTGAT-----AAT 2441
    : : : : :
QY 261 ValGluLysAspGluAlaHisValLysIleThrLysLysSerAspLeuLysAlaIleAsp 280
    : : : : :
Db 2442 GTATCTAAAAATAATAA-----TTATCACTGTCTGATATATCACTTTAATG 2489
    : : : : :
QY 281 AspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLysLeuIleAsn 300
    : : : : :
Db 2490 GATCAATTTTCGTTA-----AATCCATGTACCATAAGAAATATA-----TTATTATCT 2537
    : : : : :
QY 301 AspAspThrLysLysAspMetLeuGlyLysLeuSerThrGlyLeuValGlnAsnPhe 320
    : : : : :
Db 2538 TCAGCAACTATATAA-----TCAAAACATATTAGCGGTAGGGCAGCTAAAAAAGTGG 2588
    : : : : :
QY 321 -----ProAsnThrIleSerLysLeuIleGluGlyLysPhe---Gln 334
    : : : : :
Db 2589 AATGTTATTTCATTGCAAAATGTATCAATGTATATAAAAAATAAAGGTTGTATCGTAATG 2648
    : : : : :
QY 335 AspMetLeuAsnIleSerGlnHisGlnCysValLys-----LysGlnCysProGluAsn 352
    : : : : :
Db 2649 GATATGGTTGATTATATATCTACTACATCTTTAAATACCATAAACAATTATATGATAAA 2708
    : : : : :
QY 353 SerGlyCysPheArgHisLeuAspGluArgGluGlyCysLysCys----- 367
    : : : : :
Db 2709 ATGAGTACCTTTGAATATAAACGAGATATAAATCATGTAAATGCTCGATATGTTCCGAC 2768
    : : : : :
QY 368 ---LeuLeuAsnTyrLysGlnGluGlyAspLysCysValGluAsnProAsnProThrCys 386
    : : : : :
Db 2769 TCTATAACACATCATATATATGAAACACATCATGTATA---AATATAAATCTACCGAT 2825
    : : : : :
QY 387 AsnGlu 388
    : : : : :
Db 2826 AATGAT 2831
```

## RESULT 29

```
US-08-295-802-113
; Sequence 113, Application US/08295802
; Patent No. 6127163
; GENERAL INFORMATION:
; APPLICANT: Cochran Ph.D., Mark D
; APPLICANT: Junker M.S., David E
; TITLE OF INVENTION: Recombinant Swinepox Virus
; NUMBER OF SEQUENCES: 188
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/295,802
; FILING DATE: Herewith
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)977-9550
; TELEFAX: (212)664-0525
; TELEX: 422523
; INFORMATION FOR SEQ ID NO: 113:
```

```
;;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3628 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Swinepox virus
; STRAIN: Kasza
; INDIVIDUAL ISOLATE: S-SPV-001
; IMMEDIATE SOURCE:
; CLONE: 515-85.1
; POSITION IN GENOME:
; MAP POSITION: ~23.2
; UNITS: %G
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 57..1226
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1362..3395
;
; US-08-295-802-113
```

```
Alignment Scores:
Pred. No.: 0.00136 Length: 3628
Score: 126.00 Matches: 78
Percent Similarity: 41.81% Conservative: 65
Best Local Similarity: 22.81% Mismatches: 135
Query Match: 5.54% Indels: 64
DB: 3 Gaps: 19
```

US-10-057-531A-2 (1-431) x US-08-295-802-113 (1-3628)

```
QY 75 AspValIleTyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGlu 94
    : : : : :
Db 1914 GATGCATTCACATACCCCCACACATATATCATTCAGTTTCACCT---AGAGATATAAAC 1970
    : : : : :
QY 95 LysAsnIlePheThrPhe-----AsnLeuAsnLeuAsnAspIleLeuAsnSer 110
    : : : : :
Db 1971 AGGATTATTGAATTCGCTTAAAAAATATCCGAATAAATAATTATTGATTATATATCCGAT 2030
    : : : : :
QY 111 ArgLeuLysLysArgLysTyrPheLeuAspValLeuGlu-----SerAspLeuMet 127
    : : : : :
Db 2031 AGCATAAAATCAAAATAGTTCATTTCATTCACATATGATATGATAATCAAAATATGTTT 2090
    : : : : :
QY 128 GlnPheLysHisIleSerSerAsnGluTyrIle-----IleGluAspSerPheLysLeu 145
    : : : : :
Db 2091 CCTGCTATAATCCCTAGTGTAAACGATTTTATATCTACCGTAGTTGATAAGACGACTT 2150
    : : : : :
QY 146 LeuAsnSerGluGln---LysAsnThrLeuLeuLysSerTyrLysTyr----- 160
    : : : : :
Db 2151 ATTAAATATGATGGGATTAAGTGTGTGCTATGTTTCGTACCATATATAACATGATCGAT 2210
    : : : : :
QY 161 IleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSerTyrTyrGlu 180
    : : : : :
Db 2211 TTAGAGTCATTAGATGACTCAGATTACATATTTATAGAAAAAATAATATCTATATACGAC 2270
    : : : : :
QY 181 LysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValIleLysGluGlu 200
    : : : : :
Db 2271 -----GTTAAATGTAGAGATTTTTCGGAATATGATTAGAGATAAGGTTAAAGAGAA 2321
    : : : : :
QY 201 LysGluLysPheProSerSerProThrThrProProSerProAlaLysThrAspGlu 220
    : : : : :
Db 2322 AAGAATAGATATATACTAGC-----AAATGTGAAGAT 2354
    : : : : :
QY 221 GlnLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrAsnAsn 240
    : : : : :
Db 2355 ATTATAAGATATATAAAATTTATTCAGT-----AAAAATAGATAAAGCAT 2399
    : : : : :
QY 241 LeuValAsnLysIleAspTyrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsn 260
    : : : : :
```



```
Db 2490 GATCAATTCGGTTA-----AATCCATGATACCATAAGAAATATA-----TTATTATCT 2537
Qy 301 AspAspThrLysLysAspMetLeuGlyLysLeuSerThrGlyLeuValGlnAsnPhe 320
Db 2538 TCAGCACTATAA-----TCAAACATATTAGCGTTACGGCGCAGTAAAAAACTGG 2588
Qy 321 -----ProAsnThrIleLeuSerLysLeuGlyLysPhe---Gln 334
Db 2589 AAATGTTATTCATTCGACAAATGATCAATGATATAAAAAATAAAGGCTGTATCGTAATG 2648
Qy 335 AspMetLeuAsnIleSerGlnHisGlnCysValLys-----LysGlnCysProGluAsn 352
Db 2649 GATATGGTGGTATATATCTACTAACATCTTAAATACCATATAAACAATATATGATAAA 2708
Qy 353 SerGlyCysPheArgHisLeuAspGluArgGluGluCysLysCys----- 367
Db 2709 ATGAGTACGTTGCAATATATAACGAGATATAATCATGTAAATGCTCGATATGTTCCGAC 2768
Qy 368 ---LeuLeuAsnTyrLysGlnGluGlyAspLysCysValGluAsnProAsnProThrCys 386
Db 2769 TCTATAACACATCATATATGAAACAACATCATGTATA---AATTATAAATCTACCGAT 2825
Qy 387 AsnGlu 388
Db 2826 AATGAT 2831
RESULT 31
US-08-375-992A-113
; Sequence 113, Application US/08375992A
; Patent No. 6328975
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D.
; APPLICANT: Junker, David E.
; TITLE OF INVENTION: Recombinant Swinepox Virus
; NUMBER OF SEQUENCES: 220
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/375,992A
; FILING DATE: Herewith
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 113:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3628 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Swinepox virus
; STRAIN: Kasza
; INDIVIDUAL ISOLATE: S-SPV-001
; IMMEDIATE SOURCE:
; CLONE: 515-85.1
```

```
; POSITION IN GENOME:
; MAP POSITION: ~23.2
; UNITS: %G
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 57..1226
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1362..3395
; US-08-375-992A-113
Alignment Scores:
Pred. No.: 0.00136 Length: 3628
Score: 126.00 Matches: 78
Percent Similarity: 41.81% Conservative: 65
Best Local Similarity: 22.81% Mismatches: 135
Query Match: 5.54% Indels: 64
DB: 4 Gaps: 19
US-10-057-531A-2 (1-431) x US-08-375-992A-113 (1-3628)
Qy 75 AspValIleTyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGlu 94
Db 1914 GATGCATTCACATACCCCCACACATATATCATTCAGTTCACTT---AGAGATAAAG 1970
Qy 95 LysAsnIlePheThrPhe-----AsnLeuAsnLeuAsnAspIleLeuAsnSer 110
Db 1971 AGGATTATTCGAATTCGTTAAAAAATATCCGAATAATAATATTATGATTATATATCCGAT 2030
Qy 111 ArgLeuLysLysArgLysTyrPheLeuAspValLeuGlu-----SerAspLeuMet 127
Db 2031 AGCATAAAAATCAATATGTTTCATTCATTCACATATCATATGATATCAAAATATGTTT 2090
Qy 128 GlnPheLysHisIleSerSerAsnGluTyrIle-----IleGluAspSerPheLysLeu 145
Db 2091 CTGCTATATATCCCTAGTGTAAACGATTTTATATCATCCGATGTTGATAAAGATCGACTT 2150
Qy 146 LeuAsnSerGluGln---LysAsnThrLeuLeuLysSerTyrLysTyr----- 160
Db 2151 ATTAATATGATGGGATTAAGTGTGTTGCTATGTTTCGTACGATATAAATCATGATCGAT 2210
Qy 161 IleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSerTyrTyrGlu 180
Db 2211 TTAGAGTTCATAGATCACTCAGATATATATTTATAGAAAAAATATATATATATACGAC 2270
Qy 181 LysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValIleLysGluGlu 200
Db 2271 -----CTTAATGCTAGAGATTTTTCGAATATGATTAGAGATAAGGTTAAAGAGAA 2321
Qy 201 LysGluLysPheProSerSerProThrThrProProSerProAlaLysThrAspGlu 220
Db 2322 AAGATAGATAATATAACTACG-----AATGTGAAGAT 2354
Qy 221 GlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrAsnAsn 240
Db 2355 ATTATAAGATATATAAAATTTATTCAGT-----AAAAATAGATAAATACGAT 2399
Qy 241 LeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsn 260
Db 2400 GAAAAATAAAGTGGAGGAGGTGTTGATACATATTCAT-----AAT 2441
Qy 261 ValGluLysAspGluAlaHisValLysIleThrLysLysSerAspLeuLysAlaIleAsp 280
Db 2442 GTATCTAAAAATAATAAA-----TTATCACTGCTCATATATATCTTTTAATG 2489
Qy 281 AspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLysLeuIleAsn 300
Db 2490 GATCAATTCGTTTA-----AATCCATGTACCATAGAAATATA-----TTATTATCT 2537
Qy 301 AspAspThrLysLysAspMetLeuGlyLysLeuSerThrGlyLeuValGlnAsnPhe 320
Db 2538 TCAGCACTATAA-----TCAAAACTATTACGTTACGGCGCAGTAAAAAACTGG 2588
```

```

QY 321 -----ProAsnThrIleIleSerLysLeuIleGluGlyLysPhe---Gln 334
Db 2589 AAATGTTTATTCATTGACAAATGATCAATGATATAAAAAAATAAAGGTGTTATCGTAATG 2648
QY 335 AspMetLeuAsnIleSerGlnHisGlnCysValLys-----LysGlnCysProGluAsn 352
Db 2649 GATATGTTGATATATATCTACTAACATCTCTTAATACCATAAACAATTATATGATAAA 2708
QY 353 SerGlyCysPheArgHisLeuAspGluArgGluGlyCysLysCys----- 367
Db 2709 ATGAGTAGCTTTGAATATAAAACGAGATATAAATCATGCTAAATGCTCGATATGTTCCGAC 2768
QY 368 ---LeuLeuAsnThrLysGlnGluCysLysCysValGluAsnProAsnProThrCys 386
Db 2769 TCTATAACATCATATATATGAAACAACATCATGTATATA---AATTATAAATCTACCGAT 2825
QY 387 AsnGlu 388
Db 2826 AATGAT 2831
RESULT 32
US-08-480-640A-189
Sequence 189, Application US/08480640A
Patent No. 6033904
GENERAL INFORMATION:
APPLICANT: Cochran, Mark D.
APPLICANT: Junker, David E.
TITLE OF INVENTION: Recombinant Swinepox Virus
NUMBER OF SEQUENCES: 225
CORRESPONDENCE ADDRESS:
ADDRESSEE: John P. White
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480, 640A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: White, John P
REGISTRATION NUMBER: 28,678
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 189:
SEQUENCE CHARACTERISTICS:
LENGTH: 3942 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Swinepox virus
STRAIN: Kasza
INDIVIDUAL ISOLATE: S-SPV-001
FEATURE:
NAME/KEY: CDS
LOCATION: 1..369
FEATURE:
NAME/KEY: CDS
LOCATION: 370..597
NAME/KEY: CDS
LOCATION: 598..1539

```

```

; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1675..3708
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (3748..3942)
US-08-480-640A-189
Alignment Scores:
Pred. No.: 0.00155 Length: 3942
Score: 126.00 Matches: 78
Percent Similarity: 41.81% Conservative: 65
Best Local Similarity: 22.81% Mismatches: 135
Query Match: 5.54% Indels: 64
DB: 3 Gaps: 19
US-10-057-531A-2 (1-431) x US-08-480-640A-189 (1-3942)
QY 75 AspValIleTyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysGlnIleGlu 94
Db 2227 GATGCAATTCACATACCCCCACACATATATCATTCATTCCTCCTT---AGAGATATAAAC 2283
QY 95 LysAsnIlePheThrPhe-----AsnLeuAsnLeuAsnAspIleLeuAsnSer 110
Db 2284 AGGATTATTGAATTGCTTAAAAAATATCCGAATAATAATATATTGATTATATATCCGAT 2343
QY 111 ArgLeuLysLysArgLysTyrPheLeuAspValLeuGlu-----SerAspLeuMet 127
Db 2344 AGCATAAATCAATAGTTTCATTCATTCACATCTTCATATGATATAATATATGTTT 2403
QY 128 GlnPheLysHisIleSerSerAsnGluTyrIle-----IleGluAspSerPheLysLeu 145
Db 2404 CCTGCTATAATCCCTAGTGTAAAGGATTTTATATCTACCGTAGTTGATAAAGATCGACTT 2463
QY 146 LeuAsnSerGluGln---LysAsnThrLeuLeuLysSerTyrLysTyr----- 160
Db 2464 ATTAATATGATGGGATTAAGTGTGCTATGTTTTCGTACGATATAAAGATATCGAT 2523
QY 161 IleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSerTyrTyrGlu 180
Db 2524 TTAGAGTCATTAGATGACTCAGATACATATTTATAGAAAAAATATATCTATATACGAC 2583
QY 181 LysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValIleLysGluGlu 200
Db 2584 -----GTTAAATGTAGAGATTTTCCGAATATGATTAGAGATAAGGTTAAAGAGAA 2634
QY 201 LysGluLysPheProSerSerProProThrProProSerProAlaLysThrAspGlu 220
Db 2635 AAGAATAGAAATATTAACACTACG-----AAATGTGAAGAT 2667
QY 221 GlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrAsnAsn 240
Db 2668 ATTATAAGATATATAAAATATTTCAGT-----AAAAATAGATAAAGCAT 2712
QY 241 LeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsn 260
Db 2713 GAAATAATAAGGTGGAGAGGTTGTCATACATATTGAT-----AAT 2754
QY 261 ValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLysAlaIleAsp 280
Db 2755 GTATCTAAAAATAATAAA-----TTATCACTGCTGATATATCATCTTTAATG 2802
QY 281 AspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLysLeuIleAsn 300
Db 2803 GATCAATTTCCGTTTA-----AATCCATGTACCAATAAGAAATATA-----TTATTATCT 2850
QY 301 AspAspThrLysLysAspMetLeuGlyLysLeuSerThrGlyLeuValGlnAsnAsp 320
Db 2851 TCAGCAACTATAAAA-----TCAAAACTATTACGTTACGGCGAGTAAAAAACTGG 2901
QY 321 -----ProAsnThrIleIleSerLysLeuIleGluGlyLysPhe---Gln 334
Db 2902 AAATGTTTATTCATTGACAAATGATCAATGATATAAAAAAATAAAGGTGTTATCGTAATG 2961

```

```

QY 335 AspMetLeuAsnIleSerGlnHisGlnCysValLys-----LysGlnCysProGluAsn 352
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2962 GATATGGTTGATTATATATCTACTAATCTTAAATACCAATAAATATATATGATAA 3021
QY 353 SerGlyCysPheArgHisLeuAspGluArgGluGluCysLysCys----- 367
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3022 ATGAGTACGCTTTGCAATATATAACGAGATATTAATCATCTGTAATGCTCGATATGTTCCGAC 3081
QY 368 ---LeuLeuAsnTyrLysGlnGluGlyAspLysCysValGluAsnProAsnProThrCys 386
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3082 TCTATACACATCATATATATGAAACAACATCATGTATA---AATTATAAATCTACCGAT 3138
QY 387 AsnGlu 388
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3139 AATGAT 3144

RESULT 33
US-08-686-968C-189
; Sequence 189, Application US/08686968C
; Patent No. 6221361
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D.
; APPLICANT: Junker, David E.
; TITLE OF INVENTION: Recombinant Swinepox Virus
; FILE REFERENCE: 39119-H/JML
; CURRENT APPLICATION NUMBER: US/08/686,968C
; CURRENT FILING DATE: 1996-07-25
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 189
; LENGTH: 3942
; TYPE: DNA
; ORGANISM: Swinepox virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(369)
; NAME/KEY: CDS
; LOCATION: (370)..(597)
; NAME/KEY: CDS
; LOCATION: (598)..(1539)
; NAME/KEY: CDS
; LOCATION: (1675)..(3708)
; LOCATION: (1675)..(3708)
US-08-686-968C-189

Alignment Scores:
Pred. No.: 0.00155 Length: 3942
Score: 126.00 Matches: 78
Percent Similarity: 41.81% Conservative: 65
Best Local Similarity: 22.81% Mismatches: 135
Query Match: 5.54% Indels: 64
DB: 4 Gaps: 19

US-10-057-531A-2 (1-431) x US-08-686-968C-189 (1-3942)

QY 75 AspValIleTyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGlu 94
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2227 GATGATTCACATACCCACACATATATCATCTTACGTTCACTT---AGAGATATAAAC 2283
QY 95 LysAsnIlePheThrPhe-----AsnLeuAsnLeuAsnAspIleLeuAsnSer 110
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2284 AGGATTATTGAAATTTGCTTAAAAAATATCCGAATAATAATATTATTCATATATATCCGAT 2343
QY 111 ArgLeuLysLysArgLysTyrPheLeuAspValLeuGlu-----SerAspLeuMet 127
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2344 AGCATAAATCAATATGCTTCAATTCATTCATCATCATCTTATGATATATCAATAATGTTT 2403
QY 128 GlnPheLysHisIleSerSerAsnGluTyrIle-----IleGluAspSerPheLysLeu 145
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2404 CCGCTATAATCCCTGCTGTAACGAGTTTATATCTACCGTAGCTTGATAAAGATCGACIT 2463
QY 146 LeuAsnSerGluGln---LysAsnThrLeuLeuLysSerTyrLysTyr----- 160
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Db 2464 ATTAATATGATGGGATTAAAGTGTGTTGCTATCTTTTCTGACGATATAAATCATGATCGAT 2523
QY 161 IleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSerTyrGlu 180
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2524 TTAGAGTCATGATGACGACGAGATTACATATTTATGAAAAAATATATCTATATACGAC 2583
QY 181 LysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValIleLysGluGlu 200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2584 -----GTTAAATGTAGAGATTTTGGCAATATGATTAGACATAAGGTTAAAGAGAA 2634
QY 201 LysGluLysPheProSerSerProProThrThrProProSerProAlaLysThrAspGlu 220
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2635 AGAATAGATATTAATACG-----AAATGTGAAGAT 2667
QY 221 GlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrAsnAsn 240
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2668 ATTATAAGATATATAAATATTTCAGT-----AAAATAGATAAATACGAT 2712
QY 241 LeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsn 260
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2713 GAAATAATAAGGTGAGGAGGTTGTGATACATATTGAT-----AAT 2754
QY 261 ValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLysAlaIleAsp 280
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2755 GTATCTAAAAAATAATAA-----TTATCACTGTCTGATATATCATCTTTAATG 2802
QY 281 AspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLysLeuIleAsn 300
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2803 GATCAATTTTCGTTTA-----AATCCATGTACCATAGAATAATA-----TTATTATCT 2850
QY 301 AspAspThrLysLysAspMetLeuGlyLysLeuSerThrGlyLeuValGlnAsnPhe 320
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2851 TCAGCAACTATAAAA-----TCAAAACTATTAGCGTTACGGCGAGTAAAAAATGG 2901
QY 321 -----ProAsnThrIleIleSerLysLeuIleGluGlyLysPhe---Gln 334
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2902 AAATGTTATTTCATTGACAAATGATCAATGATATAAAAAATAAAGGTTTATCGTAAATG 2961
QY 335 AspMetLeuAsnIleSerGlnHisGlnCysValLys-----LysGlnCysProGluAsn 352
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2962 GATATGGTTGATTATATCTACTAATCTTAAATACCATCAACAATATATATGATAA 3021
QY 353 SerGlyCysPheArgHisLeuAspGluArgGluGluCysLysCys----- 367
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3022 ATGAGTACGTTTCAATATAAACGAGATATTAATCATCTGTAATGCTCGATATGTTCCGAC 3081
QY 368 ---LeuLeuAsnTyrLysGlnGluGlyAspLysCysValGluAsnProAsnProThrCys 386
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3082 TCTATAACACATCATATATATGAAACAACATCATGTATA---AATTATAAATCTACCGAT 3138
QY 387 AsnGlu 388
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3139 AATGAT 3144

RESULT 34
US-08-488-237A-189
; Sequence 189, Application US/08488237A
; Patent No. 6251403
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D.
; APPLICANT: Junker, David E.
; TITLE OF INVENTION: Recombinant Swinepox Virus
; NUMBER OF SEQUENCES: 225
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/488,237A  
 FILING DATE: 07-JUN-1995  
 CLASSIFICATION: 424  
 ATTORNEY/AGENT INFORMATION:  
 NAME: White, John P  
 REGISTRATION NUMBER: 28,678  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 278-0400  
 TELEFAX: (212) 391-0525  
 INFORMATION FOR SEQ ID NO: 189:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3942 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 ORGANISM: Swinepox virus  
 STRAIN: Kasza  
 INDIVIDUAL ISOLATE: S-SPV-001  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 1..369  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 370..597  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 598..1539  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 1675..3708  
 NAME/KEY: CDS  
 LOCATION: complement (3748..3942)

US-08-488-237A-189

Alignment Scores:  
 Pred. No.: 0.00155 Length: 3942  
 Score: 126.00 Matches: 78  
 Percent Similarity: 41.81% Conservative: 65  
 Best Local Similarity: 22.81% Mismatches: 135  
 Query Match: 5.54% Indels: 64  
 DB: 4 Gaps: 19

US-10-057-531A-2 (1-431) x US-08-488-237A-189 (1-3942)

Qy 75 AspValIleTyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGlu 94

Db 2227 GATCATTCACACACCCACACATATATCATTCATTCACCTT---AGACATATAAAC 2283

Qy 95 LysAsnIlePheThrPhe-----AsnLeuAsnLeuAsnAspIleLeuAsnSer 110

Db 2284 AGGATTATTGAATGCTTAAATAATATCCGAATATAATATATTGATTATATATCCGAT 2343

Qy 111 ArgLeuLysLysArgLysTyrPheLeuAspValLeuGlu-----SerAspLeuMet 127

Db 2344 AGCATAAATCAATAGTTCATTCATTCATCATATGATAATATCAATATGTTT 2403

Qy 128 GlnPheLysHisIleSerSerAsnGluTyrIle-----IleGluAspSerPheLysLeu 145

Db 2404 CCTGCTATAATCCCTAGTGTAAAGCGATTTTATATCTACCGTAGTTGATAAAGATCGACTT 2463

Qy 146 LeuAsnSerGluGln---LysAsnThrLeuLeuLysSerTyrLysTyr----- 160

Db 2464 ATTAAATGATGGATTAAGTGTGCTATGTTTTCGTACGATATAAACATCATCGAT 2523

Qy 161 IleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSerTyrTyrGlu 180

Db 2523

Qy 181 LysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValIleLysGluGlu 200

Db 2584 -----GTTAAATGTAGAGATTTTTCGAATATATAGAGTAAGGTTTAAAGAGAA 2634

Qy 201 LysGluLysPheProSerSerProProThrThrProProSerProAlaLysThrAspGlu 220

Db 2635 AAGATAAGATATAATAAATTTATTCAGT-----AAATGCTGAAGAT 2667

Qy 221 GlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLysThrAsnAsn 240

Db 2668 ATTATAAGATATAATAAATTTATTCAGT-----AAATGATAAATAAAGAT 2712

Qy 241 LeuValAsnIleAspAspTyrLeuIleAsnLeuLysAlaLysLysAsnAspCysAsn 260

Db 2713 GAAATAATAAGGTGGAGGAGTGTGATACATATTGAT-----AAT 2754

Qy 261 ValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLysAlaIleAsp 280

Db 2755 GTATCTAAATAATAA-----TTATCACTGTCTGATATATCATCTTTAATG 2802

Qy 281 AspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLysLeuLeasn 300

Db 2803 GATCAATTTTCGTTTA-----AATCCATGTACCATAGAAATATA-----TTATTATCT 2850

Qy 301 AspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuValGlnAsnPhe 320

Db 2851 TCAGCAACTATAA-----TCAAACTATTAGCGTTACGGCGAGTAAACATG 2901

Qy 321 -----ProAsnThrIleIleSerLysLeuIleGluGlyLysPhe---Gln 334

Db 2902 AAATGTTTATTCATTCACAAATGTATCAATGTATAAATAAAGGGTGTTCGTAATG 2961

Qy 335 AspMetLeuAsnIleSerGlnHisGlnCysValLys-----LysGlnCysProGluAsn 352

Db 2962 GATATGGTTGATTATATATCTACTAACATTTCTTAAATACCATTAACAATATATGATAA 3021

Qy 353 SerGlyCysPheArgHisLeuAspGluArgGluGluCysLysCys----- 367

Db 3022 ATGAGTACGTTTGAATATAAAGAGATATTAAATCATGTAATGCTCGATATGTTCCGAC 3081

Qy 368 ---LeuLeuAsnTyrLysGlnGluCysValGluAsnProAsnProThrCys 386

Db 3082 TCTATAACATCATATATATGAAACACATCATGTATA---AATTATAAATCTACCGAT 3138

Qy 387 AsnGlu 388

Db 3139 ATGAT 3144

RESULT 35

US-08-375-992A-189

Sequence 189, Application US/08375992A

Patent No. 6328975

GENERAL INFORMATION:

APPLICANT: Cochran, Mark D.

APPLICANT: Junker, David E.

TITLE OF INVENTION: Recombinant Swinepox Virus

NUMBER OF SEQUENCES: 220

CORRESPONDENCE ADDRESS:

ADDRESSEE: John P. White

STREET: 1185 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:



APPLICATION NUMBER: US/08/375,992A  
FILING DATE: Herewith  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P  
REGISTRATION NUMBER: 28,678  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525  
INFORMATION FOR SEQ ID NO: 189:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3942 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Swinepox virus  
STRAIN: Kasza  
INDIVIDUAL ISOLATE: S-SPV-001  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..369  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 370..597  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 598..1539  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1675..3708  
FEATURE:  
NAME/KEY: CDS  
LOCATION: complement (3748..3942)

US-08-375-992A-189

Alignment Scores:

Score: 0.00155 Length: 3942

Percent Similarity: 126.00 Matches: 78

Best Local Similarity: 41.81% Conservative: 65

Query Match: 22.81% Mismatches: 135

DB: 5.54% Indels: 64

DB: 4 Gaps: 19

US-10-057-531A-2 (1-431) x US-08-375-992A-189 (1-3942)

QY 75 AspValIleTyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGlu 94

DB 2227 GATGCATTCCACATACCCCCACACATATATCATTCAGTTCACTT--AGAGATATAAAC 2283

QY 95 LysAsnIlePheThrPhe-----AsnLeuAsnLeuAsnAspIleLeuAsnSer 110

DB 2284 AGGATTATTCAATTGCTTTAAAAAATATCCGAATAATAATATTATTGATTATATATCCGAT 2343

QY 111 ArgLeuLysArgLysTyrPheLeuAspValLeuGlu-----SerAspLeuMet 127

DB 2344 AGCATAAATCAAAATGTCATTCATTCATCATTCATCATATATATCAAAATATGTTT 2403

QY 128 GlnPheLysHisIleSerSerAsnGluTyrIle-----IleGluAspSerPheLysLeu 145

DB 2404 CCGTCTAATATCCCTAGTAGTAACGATTTTATATCTACCGTAGTTGATAAGATCGACTT 2463

QY 146 LeuAsnSerGluGln---LysAsnThrLeuLeuLysSerTyrLysTyr----- 160

DB 2464 ATTAATATGTATGGGATTAAGTGTGTGCTATGTTTCTACCATATAAATATGATCGAT 2523

QY 161 IleLysGluSerValGluAsnAspIleLysPheAlaGlnGluLysSerTyrTyrGlu 180

DB 2524 TTAGAGTCTATTAGTACTGACGATTTATTTATAGAAAAAATATATCTATATACGAC 2583

QY 181 LysValLeuAlaLysTyrLysAspLeuGluSerIleLysLysValIleLysGluGlu 200  
DB 2584 -----CTTAATCTAGAGATTTTCCGAATATGATTAGAGATAGAGTAAAGAGAA 2634  
QY 201 LysGluLysPheProSerSerProThrThrProProSerProAlaLysThrAspGlu 220  
DB 2635 AAGAATAGATATTAATACTAGC-----AAATGTTGAAGAT 2667  
QY 221 GlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrAsnAsn 240  
DB 2668 ATTATAGATATATAAATTTATTCAGT-----AAAAATAGATAAATACGAT 2712  
QY 241 LeuValAsnLysIleAspTyrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsn 260  
DB 2713 GAAAAATAAAGTGGGAGGAGGTTCATACATATTGAT-----AAT 2754  
QY 261 ValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLysAlaIleAsp 280  
DB 2755 GTATCTAAAAATAATAA-----TTATCACTGCTGATATATCATCTTTAATG 2802  
QY 281 AspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLysLeuIleAsn 300  
DB 2803 GATCAATTTGCTTTA-----AATCCATGTACCATAGAAATATA-----TTATTATCT 2850  
QY 301 AspAspThrLysLysAspMetLeuGlyLysLeuSerThrGlyLeuValGlnAsnPhe 320  
DB 2851 TCAGCAACTATAAA-----TCAAACTATTACGTTACGGCAGTAAAAAACTGG 2901  
QY 321 -----ProAsnThrIleIleSerLysLeuIleGluGlyLysPhe---Gln 334  
DB 2902 AAATGTTATTTCATTGACAAATGTATCAATGTATAAAAAATAAAGGCTGTATCGTAATG 2961  
QY 335 AspMetLeuAsnIleSerGlnHisGlnCysValLys-----LysGlnCysProGluAsn 352  
DB 2962 GATATGTTGATTATATATCTACTATAACATCTTTAAATACCATCAACAAATTATATGATAAA 3021  
QY 353 SerGlyCysPheArgHisLeuAspGluArgGluGluCysLysCys----- 367  
DB 3022 ATGAGTACGTTTGAATATAAACGAGATATTAAATCATGTAATCTCGATATGTTCCGAC 3081  
QY 368 ---LeuLeuAsnTyrLysGlnGluGlyAspLysCysValGluAsnProAsnProThrCys 386  
DB 3082 TCTATAACACATCATATATATGAACACATCATGTATA---AATTATAAATCTACCGAT 3138  
QY 387 AsnGlu 388  
DB 3139 AATGAT 3144

RESULT 36

US-08-480-640A-221

; Sequence 221, Application US/08480640A

; Patent No. 6033904

; GENERAL INFORMATION:

; APPLICANT: Cochran, Mark D.

; APPLICANT: Junker, David E.

; TITLE OF INVENTION: Recombinant Swinepox Virus

; NUMBER OF SEQUENCES: 225

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: John P. White

; STREET: 1185 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: USA

; ZIP: 10036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0; Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/480,640A

; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 424



Db 4307 ATTAATATGATGGATTAAAGTGTGCTGCTATGTTTCGTACGATATAAATCATGATCGAT 4366  
QY 161 IleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSerTyrTyrGlu 180  
Db 4367 TTAGAGTCATTAGATGACTCGAGATTACATATTTATGAAAAAATATATCTATATACGAC 4426  
QY 181 LysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValIleLysGluGlu 200  
Db 4427 -----GTTAAATGATAGAGATTTTGCGAATATGATTAGATAGAGTTAAAGAGAA 4477  
QY 201 LysGluLysPheProSerSerProProThrThrProProSerProAlaLysThrAspGlu 220  
Db 4478 AAGAATAGATATATAACTAGC-----AAATGCTGAAGAT 4510  
QY 221 GlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrAsnAsn 240  
Db 4511 ATTATAGATATATAAATATTTCAGT-----AAAAATAGAAATAACGAT 4555  
QY 241 LeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsn 260  
Db 4556 GAAAAATAAATAGCTGGAGGAGGTGTTGATACATATTGAT-----AAT 4597  
QY 261 ValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLysAlaIleAsp 280  
Db 4598 GTATCTAAAAAATAATAA-----TTATCACTGTCTGATATATCATCTTTAATG 4645  
QY 281 AspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLysLeuIleAsn 300  
Db 4646 GATCAATTCGTTTA-----AATCATGTACCATAGAATAATA-----TTATATATCT 4693  
QY 301 AspAspThrLysLysAspMetLeuGlyLysLeuSerThrGlyLeuValGlnAsnPhe 320  
Db 4694 TCAGCAACTATAAA-----TCAAAACTATTAGCGTTACGGCGAGTAAAAAATCGG 4744  
QY 321 -----ProAsnThrIleIleSerLysLeuIleGluGlyLysPhe---Gln 334  
Db 4745 AATGTTATTTCATGTGCAAAATGATCAATGTATAAAAAAATAAAGGTTGTTATCGTAATG 4804  
QY 335 AspMetLeuAsnIleSerGlnHisGlnCysValLys-----LysGlnCysProGluAsn 352  
Db 4805 GATAGTGTGATATATATCTACTAATCACTTTAAATACCATCAACAAATATATATGATAA 4864  
QY 353 SerGlyCysPheArgHisLeuAspGluArgGluGlyCysLysCys----- 367  
Db 4865 ATGAGTACGTTTGAATATAAACAGAGATATTAATCATGTAAATGCTCGATATGTTCCGAC 4924  
QY 368 ---LeuLeuAsnTyrLysGlnGluGlyAspLysCysValGluAsnProAsnProThrCys 386  
Db 4925 TCTATACATCATATATATGAAACAACATCATGTATA---AATTATAAATCTACCGAT 4981  
QY 387 AsnGlu 388  
Db 4982 AATGAT 4987

## RESULT 38

US-08-488-237A-221  
Sequence 221, Application US/08488237A  
Patent No. 6251403

## GENERAL INFORMATION:

APPLICANT: Cochran, Mark D.  
APPLICANT: Junker, David E.  
TITLE OF INVENTION: Recombinant Swinepox Virus  
NUMBER OF SEQUENCES: 225  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: John P. White  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/488,237A  
FILING DATE: 07-JUN-1995

CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:

NAME: White, John P  
REGISTRATION NUMBER: 28,678

TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525

INFORMATION FOR SEQ ID NO: 221:  
SEQUENCE CHARACTERISTICS:

LENGTH: 5785 base pairs  
TYPE: nucleic acid

STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: N

ANTI-SENSE: N  
US-08-488-237A-221

Alignment Scores:  
Pred. No.: 0.00279

Score: 126.00 Length: 5785  
Percent Similarity: 41.81% Matches: 78

Best Local Similarity: 22.81% Conservative: 65  
Query Match: 5.54% Mismatches: 135

DB: 4 Indels: 64  
Gaps: 19

US-10-057-531A-2 (1-431) x US-08-488-237A-221 (1-5785)

QY 75 AspValIleTyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysGlnIleGlu 94  
Db 4070 GATGCAATCCACATACACCCACACATATATCATTCAGTTCACTT---AGAGATATAAAC 4126  
QY 95 LysAsnIlePheThrPhe-----AsnLeuAsnLeuAsnAspIleLeuAsnSer 110  
Db 4127 AGGATTATTCATTCGTTAAAAAATATCCGAATAAATAATATATATGATATATCCGAT 4186  
QY 111 ArgLeuLysLysArgLysTyrPheLeuAspValLeuGlu-----SerAspLeuMet 127  
Db 4187 AGCATAAAATCAAAATAGTTCATTCATTCACATATCATATGATAAATAATATATATGTTT 4246  
QY 128 GlnPheLysHisIleSerSerAsnGluTyrIle-----IleGluAspSerPheLysLeu 145  
Db 4247 CCTGCTATATCCCTAGTGTAAACGATTTTATATCTACCGTAGTTGATAAAGATCGACTT 4306  
QY 146 LeuAsnSerGluGln---LysAsnThrLeuLeuLysSerTyrLysTyr----- 160  
Db 4307 ATTAATATGATGGATTAAAGTGTGCTATGTTTCGTACGATATAAATCATCGAT 4366  
QY 161 IleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSerTyrTyrGlu 180  
Db 4367 TTAGAGTCATTAGATGACTCGAGATTACATATTTATAGAAAAAATATATCTATATACGAC 4426  
QY 181 LysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValIleLysGluGlu 200  
Db 4427 -----GTTAAATGATAGAGATTTTGCGAATATGATTAGATAGAGTTAAAGAGAA 4477  
QY 201 LysGluLysPheProSerSerProProThrThrProProSerProAlaLysThrAspGlu 220  
Db 4478 AAGAATAGATATATAACTAGC-----AAATGCTGAAGAT 4510  
QY 221 GlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrAsnAsn 240  
Db 4511 ATTATAAGATATATAAATATTTCAGT-----AAAAATAGAAATAAAGAT 4555  
QY 241 LeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsn 260  
Db 4556 GAAAAATAAAGTGGAGGAGGTTGTTGATACATATTGAT-----AAT 4597

```
Qy 261 ValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLysAlaIleAsp 280
Db 4598 GTATCTAAATAATAA-----TTATCACTGTCTGATATATCATCTTTAATG 4645
Qy 281 AspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLysLeuIleAsn 300
Db 4646 GATCAATTTGGTTA-----AATCCATGTACCATAAGAAATATA-----TTATTATCT 4693
Qy 301 AspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuValClnAsnPhe 320
Db 4694 TCAGCAACTATAA-----TCAAACTATTAGGTTACGGGCAGTAAAAACTGG 4744
Qy 321 -----ProAsnThrIleIleSerLysLysLeuIleGluGlyLysPhe---Gln 334
Db 4745 AAATGTTATTCAATTCACAAATGATCAATCAATGATATAAATAAATAAAGGGTGTATCGTAATG 4804
Qy 335 AspMetLeuAsnIleSerGlnHisGlnCysValLys-----LysGlnCysProGluAsn 352
Db 4805 GATATGGTTGATTATATCTACTAACATTTCTTAATACCATAAACAATATATATGATAAA 4864
Qy 353 SerGlyCysPheArgHisLeuAspGluArgGluGluCysLysCys----- 367
Db 4865 ATGAGTACGTTTGAATATAAAGAGATATAAATCATCTGTAATGCTCGATATGTTCCGAC 4924
Qy 368 ---LeuLeuAsnTyrLysGlnGluCysLysAspLysCysValGluAsnProAsnProThrCys 386
Db 4925 TCTATAACACATCATATATATGAACAACATCATGTATA---AATTATAAATCTACCGAT 4981
Qy 387 AsnGlu 388
Db 4982 AATGAT 4987

RESULT 39
US-08-287-959-2
; Sequence 2, Application US/08287959
; Patent No. 5639651
; GENERAL INFORMATION:
; APPLICANT: Weisbach, Lawrence
; APPLICANT: Bernards, Andre
; APPLICANT: Settlement, Jeffrey
; TITLE OF INVENTION: GAP-RELATED GENE
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/287,959
; FILING DATE: August 9, 1994
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul C.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/181001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7573 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
```

```
US-08-287-959-2
Alignment Scores:
Pred. No.: 0.00479 Length: 7573
Score: 125.50 Matches: 69
Percent Similarity: 37.97% Conservative: 81
Best Local Similarity: 17.47% Mismatches: 106
Query Match: 5.52% Indels: 139
DB: 1 Gaps: 16

US-10-057-531A-2 (1-431) x US-08-287-959-2 (1-7573)
Qy 75 AspValIleTyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGlu 94
Db 2889 GAAGTTGTAAAGATTTCAGTCCCTGGCAAGGATGCACCAAGCT--CGAAGCGCTATCGA 2945
Qy 95 LysAsnIlePheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeu----- 112
Db 2946 GATCGCCTGCGAGTACTTCCGGGACCATAATAATGACATTATCAAAATCCAGGCTTTTATT 3005
Qy 113 -----LysLysArgLysTyrPheLeuAspValLeuGluSerAsp----- 125
Db 3006 CGGCGAACAAGCTCGGATGACTACAAAGACTCTCATCAATGCTGAGGATCCTCCTATG 3065
Qy 126 -----LeuMetGlnPheLysHisIle-----SerSerAsnGluTyrIleIleGluAsp 141
Db 3066 GTTGTGGTCCGAAATTTGTCCACCTGCTGGACCAAAAGTACCAGGATTTTTCAGGAGGAG 3125
Qy 142 SerPheLysLeuLeuAsnSerGlnGlnLysAsnThrLeuLeuLysSerTyrLysTyrIle 161
Db 3126 CTTGACCTTATGAAGATCGGGAAGAGGTTATCACCCCTCATTCGTCT- 3173
Qy 162 LysGluSerValGluAsnAspIleLysPhe----- 171
Db 3174 AACCAGCAGCTGGAGATGACCTCAATCTCATGGATATCAAAATTTGGACTGCTAGTGAA 3233
Qy 172 -----AlaGlnGluGlyIleSerTyrTyrGluLysValLeuAlaLysTyrLys 187
Db 3234 AATAAGATTACGTTTGCAGGATGTGTTTCCACACAGTAAAAAACTTACCAAAAAAATAAG 3293
Qy 188 AspAspLeuGlu-----SerIleLysLysValIle 197
Db 3294 GAACAGTTGCTGATATGATGATGATATAATAAACAAGAGGAGGTCTCAAGCCTTTTGAGC 3353
Qy 198 LysGluLysLysGluLysPheProSerSerPro-----ProThrThr 211
Db 3354 AAGGAGAAGAGAGAGAGACTTGGAAAGCTTACCAGCACCTGTTTATTATTATTCAAACCAAT 3413
Qy 212 ProProSerProAlaLysThrAspGlnLysLysGluSerLysPheLeuProPheLeu 231
Db 3414 CCCACCTATCTGGCCCAAGCTCATTTTCAGATGCCCAAGAACAGTCCACCAAGTTTCATG 3473
Qy 232 ThrAsn---IleGluThrLeuTyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIle 250
Db 3474 GACTCTGTAATCTTCACACTCTACAACTACCGGTCCACACAGGAGGAGTACCTGGTCTC 3533
Qy 251 -----AsnLeuLysAlaLysIleAsnAspCysAsn 260
Db 3534 CTGGCGCTCTTTAAGACAGCAGCACTCCAAGAGAAATCAAGTCGAAGGTAGATCAGATTCAA 3593
Qy 261 -----ValGluLysAspGluAlaHisValLys----- 269
Db 3594 GAGATTGTGACAGGAAATCTTACGGTTATTAAAAATGGTTGAAGTTTCAACCGGTGGTCC 3653
Qy 270 ----- 276
Db 3654 CGTGGCCAGAAATGCCCTGAGACAGATCTTGGCCCCAGTCGTGAAGAAATATATGATGAC 3713
Qy 277 LysAlaIleAsp-----AspLysIleAspLeuPheLysAsn----- 288
Db 3714 AAATCTCTCAACATCAAAACTGACCCCTGTGGATATTTACAAATCTTGGTTAATCAGATG 3773
Qy 289 -----ProTyrAspPheGluAlaIleLysLysLeu 298
```

```

Db 3774 GAGTCTCAGACAGGAGGCAAGCAAGTGCCTTATGATGTGACCCCTGAGCAGCGCTA 3833
QY 299 IleAsnAspThrLysLysAspMet----- 307
Db 3834 GCTCATGAAGAAGTGAAGACACGGCTAGACAGCTCCATCAGGAACATGCGGCTGTGACA 3893
QY 307 ----- 307
Db 3894 GACAAAGTTTCTCAGCCATTGTCAGCTCTGTGGACAAAATCCCTTATGGGATGCGCTTC 3953
QY 308 LeuGlyLysLeuLeuSerThrGlyLeuValGlnAsnPheProAsnThr----- 323
Db 3954 ATTGCCAAGTCTGAAGACTGTTGATGAGAGTTCCTTGATGCTGGTGAGGATGAG 4013
QY 324 -----IleIleSerLysLeuIleGluGlyLysPheGlnAsp 335
Db 4014 CTGCTGAAGATTATTGGTAACCTGTTTATTATTCATCATCATGAAT 4058

```

## RESULT 40

```

US-08-235-836C-73
; Sequence 73, Application US/08235836C
; Patent No. 6248562

```

## GENERAL INFORMATION:

```

; APPLICANT: Dunn, John J.
; APPLICANT: Luft, Benjamin J.
; TITLE OF INVENTION: No. 6248562el Chimeric Proteins Comprising
; TITLE OF INVENTION: Borrelia Polypeptides and Uses Therefor
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brookhaven National Laboratory
; STREET:
; CITY: Upton
; STATE: NY
; COUNTRY: USA
; ZIP: 11973

```

## COMPUTER READABLE FORM:

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/235.836C
; FILING DATE: 29-APR-1994
; CLASSIFICATION: 435

```

## PRIOR APPLICATION DATA:

```

; APPLICATION NUMBER: US 08/148,191
; FILING DATE: 01-11-93

```

## ATTORNEY/AGENT INFORMATION:

```

; NAME: Bogosian, Margaret C.
; REGISTRATION NUMBER: 25,324
; REFERENCE/DOCKET NUMBER: BNL93-28A
; TELEPHONE: (516) 282-7338
; TELEFAX: (516) 282-3729
; INFORMATION FOR SEQ ID NO: 73:

```

## SEQUENCE CHARACTERISTICS:

```

; LENGTH: 2107 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:

```

## NAME/KEY: CDS

```

; LOCATION: 1..2107

```

```

US-08-235-836C-73

```

## Alignment Scores:

```

Pred. No.: 0.000761 Length: 2107
Score: 125.00 Matches: 84
Percent Similarity: 35.19% Conservative: 61
Best Local Similarity: 20.39% Mismatches: 153
Query Match: 5.49% Indels: 114

```

```

DB: 4 Gaps: 16
US-10-057-531A-2 (1-431) x US-08-235-836C-73 (1-2107)
QY 39 AspAspAspLysLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMetAla 58
Db 262 GATCAAGATATAAAGCAAGTGTGATATTTTCTATTATTGGTAGTAAGTCAGAGCTTGAT 321
QY 59 IleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIleTyr 78
Db 322 AGTATATTAATCTAAGAAGAAATCTTACAGGG-----TAT 357
QY 79 LeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIlePhe 98
Db 358 TTAATGAAGTCTTTTATTATGAGAGGTCTAGTCGGAATTAATTCCTAAAGCTATTACA 417
QY 99 ThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyrPhe 118
Db 418 ATATATAATGCT----- 429
QY 119 LeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyrIle 138
Db 430 -----GTTTATGAGAGGAGATTAGATTATTACAAA-----GAGTTTTAT 468
QY 139 IleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSerTyr 158
Db 469 ATTGAGGCTTCTTGAAGTCTTTGACTAAAGAAATGTCAGGT---CTTTCTAGGGGTGAC 525
QY 159 -----LysTyrIleLysGlu 163
Db 526 AGTCAATGGCTGGGAAGACACAAATATTATTCTCTTTAAAGAAATATTATCTATCTGGA 585
QY 164 SerValGluAsnAspIleLysPheAlaGlnGluGlyIleSerTyrTyrGluLysValLeu 183
Db 586 AATGTTGAGTCTGACATT----- 603
QY 184 AlaLysTyrLysAspAspLeuGluSerIle-----LysLysValIle----- 197
Db 604 -----GATATTGATAGTTGGTTACAGATAAGGTGGTGCAGCTCTTTTA 648
QY 198 -----LysGluGluLysGluLysPheProSerSerProProThrThrProProSerPro 215
Db 649 AGTGAGATGAATCAGGTGTTAACTTTGCAAGAGATATTACAGACATTTCAAGCGCAACT 708
QY 216 AlaLysThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGlu 235
Db 709 CATAAAGCAGATCAAGATAAATTTGATATTGAA-----TTAGATATATTTCAT 756
QY 236 ThrLeuTyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLysLysAlaLys 255
Db 757 GAAAGTGAATCCCATATAACAGAAAACATTAGAGAATTTAAGGGATCAGCTTGAAAAGCT 816
QY 256 IleAsnAspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSer--- 274
Db 817 ACAGATGAAGAGCATAAAGAGAGATTGAAGTCAAGTGGTGTCTAAAGAAACAAAG 876
QY 275 -----AspLeuLysAlaIleAsp-----AspLysIleAspLeuPhe 286
Db 877 GAAGATTAGATAAAGGCAATTCATCTTGATAAGCTCAACAAAAATTAGATTTTGT 936
QY 287 LysAsnProTyrAspPhe-----GluAlaIleLysLysLeuIleAsn 300
Db 937 GAAGATAATCTAGATATTCAAGGGGATACTGTTAGAGAGAGAGCTTCAAGAGAATATTAC 996
QY 301 AspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuValGlnAsnPhe 320
Db 997 GAGACTAATAAGGAAAGAAATTTACCAAG----- 1026
QY 321 ProAsnThrIleIleSerLysLeuIleGluGlyLysPheGln-----AspMetLeuAsn 338
Db 1027 CCTGGTGATGTAAAGTTCCTAAAGTTGATAGCAATACAAATAAAGAGCGCTGGAA 1086
QY 339 IleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArgHis 358

```

```
Db 1087 GATTGCAGGACGAG---CTTAAAGAACTGGTGATGAAAT----- 1125
QY 359 LeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLysCys 378
Db 1126 -----CAGAAAAGAGAAATTGAAAGCAATTGCAATCAAAAAAGTGATGAAAGCCTT 1179
QY 379 ValGluAsnProAsnProThrCysAsnGluAsnGlyGlyCysAspAlaAspAlaThr 398
Db 1180 TTAATAAGTAAAGATGATAAGCAAGTAAAGATGCTAAAGCCTTGGATCTTGTATCGAGAA 1239
QY 399 CysThrGluGluAspSerGlySerSerArgLysLys 410
Db 1240 TTAATTCTAAAGCTTCTAGCAAGAAAAAGTAAA 1275
```

Search completed: May 19, 2003, 16:17:20  
Job time : 139 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 10:16:12 : Search time 31 seconds  
(without alignments)  
409.074 Million cell updates/sec

Title: US-10-057-531A-2  
Perfect score: 2275  
Sequence: 1 MHHHHHSSGLVPRGSMKE.....TCECTKPSYPLFDGIFCSS 431

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 2942292 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/1aa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1931	84.9	394	4	US-08-195-705-2
2	1921	84.4	394	4	US-08-195-705-4
3	1040	45.7	377	4	US-08-195-705-5
4	986	43.3	375	4	US-08-195-705-3
5	288	12.7	53	1	US-08-290-919-4
6	278.5	12.2	824	4	US-09-626-589-3
7	277	12.2	48	1	US-08-290-919-12
8	275	12.1	467	4	US-09-513-442-2
9	274	12.0	48	1	US-08-290-919-2
10	274	12.0	167	4	US-09-741-243C-4
11	274	12.0	551	4	US-09-741-243C-2
12	272	12.0	53	1	US-08-290-919-3
13	271	11.9	48	1	US-08-290-919-1
14	259.5	11.4	408	2	US-08-683-007A-2
15	256.5	11.3	106	1	US-08-290-919-11
16	249.5	11.0	209	4	US-09-166-966E-11
17	235	10.3	159	4	US-09-166-966E-8
18	173.5	7.6	382	4	US-09-277-716-22
19	173.5	7.6	382	4	US-09-609-161B-22
20	152.5	6.7	288	4	US-09-273-839A-8
21	151	6.6	711	3	US-08-946-475-9
22	151	6.6	711	3	US-09-340-479-9
23	139.5	6.1	1010	4	US-09-134-001C-5178
24	133	5.8	10182	4	US-09-134-001C-3159
25	128	5.6	1038	4	US-09-541-782-4
26	128	5.6	1038	4	US-09-723-820-4
27	127.5	5.6	849	4	US-09-157-257-4

28	126.5	5.6	878	4	US-09-134-001C-4378	Sequence 4378, Ap
29	126.5	5.6	1494	3	US-08-755-587-186	Sequence 186, App
30	126	5.5	677	3	US-08-480-640A-115	Sequence 115, App
31	126	5.5	677	3	US-08-480-640A-193	Sequence 193, App
32	126	5.5	677	3	US-08-295-802-115	Sequence 115, App
33	126	5.5	677	4	US-08-686-968C-58	Sequence 58, Appl
34	126	5.5	677	4	US-08-686-968C-193	Sequence 193, App
35	126	5.5	677	4	US-08-488-237A-115	Sequence 115, App
36	126	5.5	677	4	US-08-488-237A-193	Sequence 193, App
37	126	5.5	677	4	US-08-375-992A-115	Sequence 115, App
38	126	5.5	677	4	US-08-375-992A-193	Sequence 193, App
39	125.5	5.5	1657	1	US-08-287-959-1	Sequence 1, Appl
40	125	5.5	700	4	US-08-235-836C-74	Sequence 74, Appl
41	122.5	5.4	2710	1	US-08-480-604A-6	Sequence 6, Appl
42	122.5	5.4	2710	2	US-08-405-496A-6	Sequence 6, Appl
43	122.5	5.4	2710	4	US-08-915-136-6	Sequence 6, Appl
44	122.5	5.4	2710	4	US-08-957-310-6	Sequence 6, Appl
45	119.5	5.3	2285	4	US-09-308-373-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1  
US-08-195-705-2  
; Sequence 2, Application US/08195705  
; Patent No. 6420523  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Sandra  
; APPLICANT: Hui, George  
; APPLICANT: Barr, Philip  
; APPLICANT: Gibson, Helen  
; TITLE OF INVENTION: BACULOVIRUS PRODUCED PLASMIDIUM  
; TITLE OF INVENTION: FALCIPARUM VACCINE  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSES: Davis Hoxie Faithfull Hapgood  
; STREET: 45 Rockefeller Pl.  
; CITY: New York  
; STATE: N.Y.  
; COUNTRY: USA  
; ZIP: 10111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/195,705  
; FILING DATE: 14-FEB-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jacobs, Seth H  
; REGISTRATION NUMBER: 32140  
; REFERENCE/DOCKET NUMBER: 11880A3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-757-2200  
; TELEFAX: 212-586-1461  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 394 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: C-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: Plasmodium falciparum  
; STRAIN: falciparum uganda palo alto (FUP)  
; US-08-195-705-2

Query Match 84.4%; Score 1931; DB 4; Length 394;  
Best Local Similarity 98.7%; Pred. No. 2.4e-137;  
Matches 369; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 58 AISVTMDNILSGFENEYDVYILKPLAGVYRSLLKQIEKNIFTFNLNDILNSRLKRRY 117  
DB 1 AISVTMDNILSGFENEYDVYILKPLAGVYRSLLKQIEKNIFTFNLNDILNSRLKRRY 60

QY 118 FLDVLESOLMQPKHSSNEYIIEESFKLLNSQKNTLLKSYKIKESVENDIKFAQEGIS 177  
DB 61 FLDVLESOLMQPKHSSNEYIIEESFKLLNSQKNTLLKSYKIKESVENDIKFAQEGIS 120

QY 178 YVEKVLAKYKDDLESIKKVIKEEKEKFPSSPTTTPPSAKTDEQKESKFLPFLNIETL 237  
DB 121 YVEKVLAKYKDDLESIKKVIKEEKEKFPSSPTTTPPSAKTDEQKESKFLPFLNIETL 180

QY 238 YNNLVNKIDDDYILNLKAKINDCNVEKDEAHVKITKLSOLKAIDDKIDLFKNPYDFAIKK 297  
DB 181 YNNLVNKIDDDYILNLKAKINDCNVEKDEAHVKITKLSOLKAIDDKIDLFKNHDEFAIKK 240

QY 298 LINDDTKMDLGLKLLSTGLVQNFNTIISKLEGFQDMLNISQHCYVKKQCPENSGGCFR 357  
DB 241 LINDDTKMDLGLKLLSTGLVQNFNTIISKLEGFQDMLNISQHCYVKKQCPENSGGCFR 300

QY 358 HLDRECECKLLNYKQEGDKCVENPNPTCNENNGCGDADATCTEEDSGSSRKKITCECTK 417  
DB 301 HLDRECECKLLNYKQEGDKCVENPNPTCNENNGCGDADAKTCTEEDSGSNGKKITCECTK 360

QY 418 PDSYPLFDGIFCSS 431  
DB 361 PDSYPLFDGIFCSS 374

RESULT 2  
US-08-195-705-4  
; Sequence 4, Application US/08195705  
; Patent No. 6420523  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Sandra  
; APPLICANT: Hui, George  
; APPLICANT: Barr, Philip  
; APPLICANT: Gibson, Helen  
; TITLE OF INVENTION: BACULOVIRUS PRODUCED PLASMODIUM  
; TITLE OF INVENTION: FALCIPARUM VACCINE  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Davis Hoxie Faithfull Hapgood  
; STREET: 45 Rockefeller Pl.  
; CITY: New York  
; STATE: N.Y.  
; COUNTRY: USA  
; ZIP: 10111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/195,705  
; FILING DATE: 14-FEB-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jacobs, Seth H  
; REGISTRATION NUMBER: 32140  
; REFERENCE/DOCKET NUMBER: 11880A3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-757-2200  
; TELEFAX: 212-586-1461  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 394 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single

TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
FRAGMENT TYPE: C-terminal  
ORIGINAL SOURCE:  
ORGANISM: Plasmodium falciparum  
STRAIN: MAD  
US-08-195-705-4

Query Match 84.4%; Score 1921; DB 4; Length 394;  
Best Local Similarity 98.4%; Pred. No. 1.4e-136;  
Matches 368; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 58 AISVTMDNILSGFENEYDVYILKPLAGVYRSLLKQIEKNIFTFNLNDILNSRLKRRY 117  
DB 1 AISVTMDNILSGFENEYDVYILKPLAGVYRSLLKQIEKNIFTFNLNDILNSRLKRRY 60

QY 118 FLDVLESOLMQPKHSSNEYIIEESFKLLNSQKNTLLKSYKIKESVENDIKFAQEGIS 177  
DB 61 FLDVLESOLMQPKHSSNEYIIEESFKLLNSQKNTLLKSYKIKESVENDIKFAQEGIS 120

QY 178 YVEKVLAKYKDDLESIKKVIKEEKEKFPSSPTTTPPSAKTDEQKESKFLPFLNIETL 237  
DB 121 YVEKVLAKYKDDLESIKKVIKEEKEKFPSSPTTTPPSAKTDEQKESKFLPFLNIETL 180

QY 238 YNNLVNKIDDDYILNLKAKINDCNVEKDEAHVKITKLSOLKAIDDKIDLFKNPYDFAIKK 297  
DB 181 YNNLVNKIDDDYILNLKAKINDCNVEKDEAHVKITKLSOLKAIDDKIDLFKNHDEFAIKK 240

QY 298 LINDDTKMDLGLKLLSTGLVQNFNTIISKLEGFQDMLNISQHCYVKKQCPENSGGCFR 357  
DB 241 LINDDTKMDLGLKLLSTGLVQNFNTIISKLEGFQDMLNISQHCYVKKQCPENSGGCFR 300

QY 358 HLDRECECKLLNYKQEGDKCVENPNPTCNENNGCGDADATCTEEDSGSSRKKITCECTK 417  
DB 301 HLDRECECKLLNYKQEGDKCVENPNPTCNENNGCGDADATCTEEDSGSRRKKITCECTK 360

QY 418 PDSYPLFDGIFCSS 431  
DB 361 PDSYPLFDGIFCSS 374

RESULT 3  
US-08-195-705-5  
; Sequence 5, Application US/08195705  
; Patent No. 6420523  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Sandra  
; APPLICANT: Hui, George  
; APPLICANT: Barr, Philip  
; APPLICANT: Gibson, Helen  
; TITLE OF INVENTION: BACULOVIRUS PRODUCED PLASMODIUM  
; TITLE OF INVENTION: FALCIPARUM VACCINE  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Davis Hoxie Faithfull Hapgood  
; STREET: 45 Rockefeller Pl.  
; CITY: New York  
; STATE: N.Y.  
; COUNTRY: USA  
; ZIP: 10111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/195,705  
; FILING DATE: 14-FEB-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jacobs, Seth H  
; REGISTRATION NUMBER: 32140



19-290-919-4  
 uence 4, Application US/082909019  
 uent No. 5720959  
 NERAL INFORMATION:  
 APPLICANT: HOLDER, ANTHONY A.  
 APPLICANT: BLACKMAN, MICHAEL J.  
 APPLICANT: CHAPPEL, JONATHAN A.  
 TITLE OF INVENTION: IMPROVEMENT  
 TITLE OF INVENTION: VACCINE  
 NUMBER OF SEQUENCES: 19  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: CUSHMAN DABRY & C

```
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,919
FILING DATE: 04-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9203821.5
FILING DATE: 22-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB93/00367
FILING DATE: 22-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N.
REGISTRATION NUMBER: 16,773
REFERENCE/DOCKET NUMBER: 212242/HCM/MJL/6BC8/
TELEPHONE: (202) 861-3000
TELEFAX: (202) 822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 53 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /label= X
OTHER INFORMATION: /note= "X" = M and N, or "N"
US-08-290-919-4

Query Match 12.7%; Score 288; DB 1; Length 53;
Best Local Similarity 100.0%; Pred. No. 3.7e-15;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 382 PNPTCNNGGCDATCTEEDSGSRKKTCTCKPDSYPLFDGIFCSS 431
|||||
Db 2 PNPTCNNGGCDATCTEEDSGSRKKTCTCKPDSYPLFDGIFCSS 51
|||||

RESULT 6
US-09-626-589-3
Sequence 3, Application US/09626589
Patent No. 6326164
GENERAL INFORMATION:
APPLICANT: Rice, John
APPLICANT: Klotz, Andreas
APPLICANT: Crawford, John
APPLICANT: Lanning, Beth
APPLICANT: Stewart, Sandy
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
IDENTIFICATION OF MODULATORS OF DEOXYXYLULOSE 5-PHOSPHATE
TITLE OF INVENTION: SYNTHASE ACTIVITY
FILE REFERENCE: 2037 US
CURRENT APPLICATION NUMBER: US/09/626,589
CURRENT FILING DATE: 2000-07-27
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 824
TYPE: PRT
ORGANISM: Artificial Sequence
```

```
FEATURE:
OTHER INFORMATION: Residues 1-165 are from the thioredoxin sequence
OTHER INFORMATION: found in the vector pET32 supplied by No. 6326164aagen.
OTHER INFORMATION: Residues 166-824 represent the tDPS sequence from
OTHER INFORMATION: Arabidopsis shown in SEQ ID NO:2.
US-09-626-589-3

Query Match 12.2%; Score 278.5; DB 4; Length 824;
Best Local Similarity 37.4%; Pred. No. 6.9e-13;
Matches 76; Conservative 25; Mismatches 49; Indels 53; Gaps 7;

Qy 1 MHHHHSSGLVPRGSGMKETAARFERQHMDSPDLGTDGDDDDKAMADIGSIEGRGTMAIS 60
|||||
Db 116 MHHHHSSGLVPRGSGMKETAARFERQHMDSPDLGTDGDDDDKAMADIGSAS----- 167
|||||

Qy 61 VTMDNLSGFENEVDVYLKPLA--GYRSLLKQIEKNIETFNLDNLSRLKRRKYF 118
|||||
Db 168 -----LAEKGEYS-NRPPTPLDITINYPPIH-MKNLSVKELQL 204
|||||

Qy 119 LDVLESDLMQFK-----HISSNEYIIEFSKLLNSEQNTLLKSKYKIKESVENDIKFA 172
|||||
Db 205 SDELRSQVI-FNVSKTGGHGLSSLGVE-----LTVALHYIENPQDKILWD 250
|||||

Qy 173 QEGISYYEKVLAQYKDDLESIKK 195
|||||
Db 251 VGHQSYPHKILTRRRKMPTRQ 273
|||||

RESULT 7
US-08-290-919-12
Sequence 12, Application US/08290919
Patent No. 5720959
GENERAL INFORMATION:
APPLICANT: HOLDER, ANTHONY A.
APPLICANT: BLACKMAN, MICHAEL J.
APPLICANT: CHAPPEL, JONATHAN A.
TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA
TITLE OF INVENTION: VACCINE
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DABY & CUSHMAN, L.L.P.
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,919
FILING DATE: 04-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9203821.5
FILING DATE: 22-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB93/00367
FILING DATE: 22-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N.
REGISTRATION NUMBER: 16,773
REFERENCE/DOCKET NUMBER: 212242/HCM/MJL/6BC8/
TELEPHONE: (202) 861-3000
TELEFAX: (202) 822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 48 amino acids
TYPE: amino acid
```

STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-290-919-12

Query Match 12.2%; Score 277; DB 1; Length 48;  
Best Local Similarity 97.9%; Pred. No. 2.1e-14;  
Matches 47; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 338 NISOHCVKKQCPENSGCFRHLDERECKLLNYKQEGDKCVENPNT 385  
|||||  
DB 1 NISOHCVKKQCPENSGCFRHLDERECKLLNYKQEGDKCVENPNT 48

## RESULT 8

US-09-513-442-2  
Sequence 2, Application US/09513442

Patent No. 6387664

GENERAL INFORMATION:

APPLICANT: Ikemoto, Mitsushi

TITLE OF INVENTION: SPARC Fusion Protein and a Process for Producing the

TITLE OF INVENTION: Same

FILE REFERENCE: HIRAKI-04218

CURRENT APPLICATION NUMBER: US/09/513,442

CURRENT FILING DATE: 2000-02-25

NUMBER OF SEQ ID NOS: 4

SOFTWARE: Patent In Ver. 2.0

SEQ ID NO 2

LENGTH: 467

TYPE: PRT

ORGANISM: Mus musculus

US-09-513-442-2

Query Match 12.1%; Score 275; DB 4; Length 467;  
Best Local Similarity 98.0%; Pred. No. 6e-13;  
Matches 50; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHSSGLVPRGSGMKETAATAKFERQHMDSPDLGTDKAMADIGSI 51  
|||||  
DB 116 MHHHHSSGLVPRGSGMKETAATAKFERQHMDSPDLGTDKAMADIGSM 166

## RESULT 9

US-08-290-919-2

Sequence 2, Application US/08290919

Patent No. 5720959

GENERAL INFORMATION:

APPLICANT: HOLDER, ANTHONY A.

APPLICANT: BLACKMAN, MICHAEL J.

APPLICANT: CHAPPEL, JONATHAN A.

TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA

TITLE OF INVENTION: VACCINE

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.

STREET: 1100 NEW YORK AVENUE, N.W.

CITY: WASHINGTON

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20005-3918

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/290,919

FILING DATE: 04-OCT-1994

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9203821.5

FILING DATE: 22-FEB-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/GB93/00367  
FILING DATE: 22-FEB-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: KOKULIS, PAUL N.  
REGISTRATION NUMBER: 16,773  
REFERENCE/DOCKET NUMBER: 212242/HCM/MJL/6BC/8/  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 861-3000  
TELEFAX: (202) 822-0944  
TELEX: 6714627 CUSH

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 48 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

FEATURE:

NAME/KEY: Modified-site

LOCATION: 1

OTHER INFORMATION: /label= X

OTHER INFORMATION: /note= "X" = M and N, or N"

US-08-290-919-2

Query Match 12.0%; Score 274; DB 1; Length 48;

Best Local Similarity 100.0%; Pred. No. 3.6e-14;

Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 339 ISOHCVKKQCPENSGCFRHLDERECKLLNYKQEGDKCVENPNT 385

|||||

DB 2 ISOHCVKKQCPENSGCFRHLDERECKLLNYKQEGDKCVENPNT 48

RESULT 10

US-09-741-243C-4

Sequence 4, Application US/09741243C

Patent No. 6399352

GENERAL INFORMATION:

APPLICANT: Crawford Jr., John Milton

APPLICANT: Rice, John

APPLICANT: Sevala, Veeresh

APPLICANT: Stewart, Sandy

TITLE OF INVENTION: NOVEL PLANT PORPHOBILINOGEN AND FUSION

TITLE OF INVENTION: PROTEIN THEREOF

FILE REFERENCE: 20220S

CURRENT APPLICATION NUMBER: US/09/741,243C

CURRENT FILING DATE: 2000-12-19

PRIOR APPLICATION NUMBER: 60/171,785

PRIOR FILING DATE: 1999-12-22

NUMBER OF SEQ ID NOS: 8

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 4

LENGTH: 167

TYPE: PRT

ORGANISM: Unknown

FEATURE:

OTHER INFORMATION: Thioredoxin functional fragment

US-09-741-243C-4

Query Match 12.0%; Score 274; DB 4; Length 167;

Best Local Similarity 100.0%; Pred. No. 1.8e-13;

Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHSSGLVPRGSGMKETAATAKFERQHMDSPDLGTDKAMADIGS 50

|||||

DB 116 MHHHHSSGLVPRGSGMKETAATAKFERQHMDSPDLGTDKAMADIGS 165

RESULT 11

US-09-741-243C-2

Sequence 2, Application US/09741243C

Patent No. 6399352

GENERAL INFORMATION:

APPLICANT: Crawford Jr., John Milton  
APPLICANT: Rice, John  
APPLICANT: Sevala, Veeresh  
APPLICANT: Stewart, Sandy  
TITLE OF INVENTION: NOVEL PLANT PORPHOBILINOGEN AND FUSION  
TITLE OF INVENTION: PROTEIN THEREOF  
FILE REFERENCE: 2022US  
CURRENT APPLICATION NUMBER: US/09/741.243C  
CURRENT FILING DATE: 2000-12-19  
PRIOR APPLICATION NUMBER: 60/171,785  
PRIOR FILING DATE: 1999-12-22  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 551  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: A plant thioredoxin-porphobilinogen  
OTHER INFORMATION: synthase fusion protein  
US-09-741-243C-2

Query Match 12.0%; Score 274; DB 4; Length 551;  
Best Local Similarity 100.0%; Pred. No. 8.8e-13;  
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHHHHHSSGLVPRSGMKETAAAKFERQHMDSPDLGTDGDDKAMADIGS 50  
Db 116 MHHHHHSSGLVPRSGMKETAAAKFERQHMDSPDLGTDGDDKAMADIGS 165

## RESULT 12

US-08-290-919-3  
Sequence 3, Application US/08290919  
Patent No. 5720959  
GENERAL INFORMATION:  
APPLICANT: HOLDER, ANTHONY A.  
APPLICANT: BLACKMAN, MICHAEL J.  
APPLICANT: CHAPPEL, JONATHAN A.  
TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA  
TITLE OF INVENTION: VACCINE  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.  
STREET: 1100 NEW YORK AVENUE, N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005-3918  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/290.919  
FILING DATE: 04-OCT-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9203821.5  
FILING DATE: 22-FEB-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB93/00367  
FILING DATE: 22-FEB-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: KOKULIS, PAUL N.  
REGISTRATION NUMBER: 16,773  
REFERENCE/DOCKET NUMBER: 212242/HCM/MJL/6BC8/  
TELEPHONE: (202) 861-3000  
TELEFAX: (202) 822-0944  
INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:  
LENGTH: 53 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1  
OTHER INFORMATION: /label-X  
OTHER INFORMATION: /note="X" - M and N, or N"  
US-08-290-919-3

Query Match 12.0%; Score 272; DB 1; Length 53;  
Best Local Similarity 94.0%; Pred. No. 5.8e-14;  
Matches 47; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 382 PNPTCNENNGCDATCTEEDSGSRKKITCECTKPDSPFLDFGICSS 431  
Db 2 PNPTCNENNGCDADAKCTEEDSGSGNGKKTCECTKPDSPFLDFGICSS 51

## RESULT 13

US-08-290-919-1  
Sequence 1, Application US/08290919  
Patent No. 5720959  
GENERAL INFORMATION:  
APPLICANT: HOLDER, ANTHONY A.  
APPLICANT: BLACKMAN, MICHAEL J.  
APPLICANT: CHAPPEL, JONATHAN A.  
TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA  
TITLE OF INVENTION: VACCINE  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.  
STREET: 1100 NEW YORK AVENUE, N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005-3918  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/290.919  
FILING DATE: 04-OCT-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9203821.5  
FILING DATE: 22-FEB-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB93/00367  
FILING DATE: 22-FEB-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: KOKULIS, PAUL N.  
REGISTRATION NUMBER: 16,773  
REFERENCE/DOCKET NUMBER: 212242/HCM/MJL/6BC8/  
TELEPHONE: (202) 861-3000  
TELEFAX: (202) 822-0944  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 48 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1

OTHER INFORMATION: /label= X  
OTHER INFORMATION: /note= "X" - M and N, or "N"  
US-08-290-919-1

Query Match 11.9%; Score 271; DB 1; Length 48;  
Best Local Similarity 97.9%; Pred. No. 6.1e-14;  
Matches 46; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 339 ISQHCVKKQCPNSGCFRHLDERECKCLLNKQEGDKCVENPNPT 385  
|||||  
DB 2 ISQHCVKKQCPNSGCFRHLDERECKCLLNKQEGDKCVENPNPT 48

RESULT 14

US-08-683-007A-2

; Sequence 2, Application US/08683007A

; Patent No. 5858724

; GENERAL INFORMATION:

; APPLICANT: No. 5858724y, Robert E

; APPLICANT: Domiano, Michael

; APPLICANT: Yaeger, Keith

; APPLICANT: Kroecker, Warren

; TITLE OF INVENTION: Recombinant Rabbit Tissue Factor

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Quarles & Brady

; STREET: 1 South Pinckney Street

; CITY: Madison

; STATE: WI

; COUNTRY: USA

; ZIP: 53703

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent in Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; FILING DATE: 16-JUL-1996

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Seay, Nicholas J

; REGISTRATION NUMBER: 27386

; REFERENCE/DOCKET NUMBER: 740380.90040

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 608-251-5000

; TELEFAX: 608-251-9166

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 408 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-683-007A-2

Query Match 11.4%; Score 259.5; DB 2; Length 408;

Best Local Similarity 32.0%; Pred. No. 7.3e-12;

Matches 77; Conservative 19; Mismatches 56; Indels 89; Gaps 9;

QY 1 MHHHHHSSGLVPRGSGMKETAAPERQHMDSPDLGTDGDDKAMA-----DIGSIEGRG 55

|||||

DB 116 MHHHHHSSGLVPRGSGMKETAAPERQHMDSPDLGTDGDDKAMAISDPLAIDTGR- 174

|||||

QY 56 TMAISVTMDNILSGFENEVDIVLKLPLAGVYRSLKQIEKNFTFNILNLDJLSRLKKR 115

|||||

DB 175 -----AYNLTW-----KS 182

QY 116 KYFLDVLESLDMOKFHISSNEYIIEPSKLLNSEQNTLLKSYK-IKESVENDIKFAQE 174

|||||

DB 183 TNFKTILEWPKSIDHV-----YTVQISTRLNWKSKCFLTAETECDLTDEVVKDV----- 233

|||||

QY 175 GISYIEKVLAKYKDDLESIKKVIKEEKFPSPPTTPPSPAKTDQKESKFLPEL-TN 233

|||||

DB 234 GOTYMARVLS-----YPARGNTTGFPEBPPF-----RNSPEETPYLDTN 273

QY 234 I 234

DB 274 L 274

RESULT 15

US-08-290-919-11

; Sequence 11, Application US/08290919

; Patent No. 5720959

; GENERAL INFORMATION:

; APPLICANT: HOLDER, ANTHONY A.

; APPLICANT: BLACKMAN, MICHAEL J.

; APPLICANT: CHAPPEL, JONATHAN A.

; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA

; TITLE OF INVENTION: VACCINE

; NUMBER OF SEQUENCES: 19

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.

; STREET: 1100 NEW YORK AVENUE, N.W.

; CITY: WASHINGTON

; STATE: D.C.

; COUNTRY: U.S.A.

; ZIP: 20005-3918

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent in Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; FILING DATE: 04-OCT-1994

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: GB 9203821.5

; FILING DATE: 22-FEB-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/GB93/00367

; FILING DATE: 22-FEB-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: KOKULIS, PAUL N.

; REGISTRATION NUMBER: 16,773

; REFERENCE/DOCKET NUMBER: 212242/HCM/MJL/6BC8/

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 861-3000

; TELEFAX: (202) 822-0944

; TELEX: 6714827 CUSH

; INFORMATION FOR SEQ ID NO: 11:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 106 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-290-919-11

Query Match 11.3%; Score 256.5; DB 1; Length 106;

Best Local Similarity 48.0%; Pred. No. 2.1e-12;

Matches 48; Conservative 18; Mismatches 31; Indels 3; Gaps 3;

QY 335 DMLNIS-QHCYV-KKQCPNSGCFRHLDERECKCLLNKQ-EGDKCVENPNPTCENNNG 391

|||||

DB 5 DLLGVDPKHVCVTRDIPKNAGCFRDNNGTEWRCLLYGKKGEGNTCVNNPTCDLNG 64

|||||

QY 392 GCADATCTEDSGSSRKKITCECTKPDSPDLFDGIFCSS 431

|||||

DB 65 GCOPTASQNAESTENSKKIICICTKEPTPNAYVEGVFCSS 104

|||||

RESULT 16

US-09-166-966E-11

; Sequence 11, Application US/09166966E

; Patent No. 6407208  
; GENERAL INFORMATION:  
; APPLICANT: CHEN, DAVID CHANHAN  
; APPLICANT: HU, NIEN-TAI  
; APPLICANT: CHEN, YUN-JU  
; APPLICANT: HSEU, TZONG-HSIUNG  
; TITLE OF INVENTION: CHIMERIC PROTEINS WITH A CELLULOSE BINDING DOMAIN  
; FILE REFERENCE: 32350-150960  
; CURRENT APPLICATION NUMBER: US/09/166.966E  
; CURRENT FILING DATE: 1998-10-06  
; PRIOR APPLICATION NUMBER: TW 86114750  
; PRIOR FILING DATE: 1997-10-08  
; NUMBER OF SEQ ID NOS: 11  
; SEQ ID NO 11  
; LENGTH: 209  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; NAME/KEY: thioredoxin  
; LOCATION: 1..109  
; NAME/KEY: Cellulose-binding domain  
; LOCATION: 162..197  
; NAME/KEY: RGD  
; LOCATION: 204..206  
; OTHER INFORMATION: Combined amino acid sequence of thioredoxin,  
; artificial sequence, cellulose-binding domain, and  
; OTHER INFORMATION: artificial sequence; the two artificial sequences  
; OTHER INFORMATION: flanking the cellulose-binding domain contain restriction sites  
US-09-166-966E-11

Query Match 11.0%; Score 249.5; DB 4; Length 209;  
Best Local Similarity 65.0%; Pred. No. 1.7e-11;  
Matches 52; Conservative 6; Mismatches 15; Indels 7; Gaps 2;  
QY 1 MHHHHHSSGLVPRGSGMKETAATAAKFERQHMDSPDLGTDGDDDKA-----DIGSIEGR 54  
DB 116 MHHHHHSSGLVPRGSGMLETAAAKFERQHMDSPDLGTDGDDDKANATOSHWCQCGGIGYS 175  
QY 55 G-TAIVSVMNLSGFENE 73  
DB 176 GPTVCASGTTCCVLNPNYSQ 195

RESULT 17  
US-09-166-966E-8  
; Sequence 8, Application US/09166966E  
; Patent No. 6407208  
; GENERAL INFORMATION:  
; APPLICANT: CHEN, DAVID CHANHAN  
; APPLICANT: HU, NIEN-TAI  
; APPLICANT: CHEN, YUN-JU  
; APPLICANT: HSEU, TZONG-HSIUNG  
; TITLE OF INVENTION: CHIMERIC PROTEINS WITH A CELLULOSE BINDING DOMAIN  
; FILE REFERENCE: 32350-150960  
; CURRENT APPLICATION NUMBER: US/09/166.966E  
; CURRENT FILING DATE: 1998-10-06  
; PRIOR APPLICATION NUMBER: TW 86114750  
; PRIOR FILING DATE: 1997-10-08  
; NUMBER OF SEQ ID NOS: 11  
; SEQ ID NO 8  
; LENGTH: 159  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; NAME/KEY: thioredoxin  
; LOCATION: 1..109  
; OTHER INFORMATION: Combined amino acid of thioredoxin and an artificial sequence  
US-09-166-966E-8

Query Match 10.3%; Score 235; DB 4; Length 159;  
Best Local Similarity 95.5%; Pred. No. 1.5e-10;  
Matches 42; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MHHHHHSSGLVPRGSGMKETAATAAKFERQHMDSPDLGTDGDDDKA 44  
DB 116 MHHHHHSSGLVPRGSGMLETAAAKFERQHMDSPDLGTDGDDDKS 159

RESULT 18  
US-09-277-716-22  
; Sequence 22, Application US/09277716A  
; Patent No. 6232107  
; GENERAL INFORMATION:  
; APPLICANT: Bryan, Bruce  
; APPLICANT: Szent-Gyorgyi, Christopher  
; APPLICANT: PROLUME, LTD.  
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE  
; CURRENT APPLICATION NUMBER: US/09/277.716A  
; CURRENT FILING DATE: 1999-03-26  
; EARLIER APPLICATION NUMBER: 60/102,939  
; EARLIER FILING DATE: 1998-10-01  
; EARLIER APPLICATION NUMBER: 60/089,367  
; EARLIER FILING DATE: 1998-06-15  
; EARLIER APPLICATION NUMBER: 60/079,624  
; EARLIER FILING DATE: 1998-03-27  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 22  
; LENGTH: 382  
; TYPE: PRT  
; ORGANISM: Artificial Sequence: fusion protein  
; FEATURE:  
; OTHER INFORMATION: Cellulose Binding Domain (CBD)-Gaussia luciferase fusion prote  
US-09-277-716-22

Query Match 7.6%; Score 173.5; DB 4; Length 382;  
Best Local Similarity 89.7%; Pred. No. 1.9e-05;  
Matches 35; Conservative 1; Mismatches 0; Indels 3; Gaps 1;  
QY 8 SSGLVPRGS---GMKETAATAAKFERQHMDSPDLGTDGDDDK 43  
DB 159 SAGLVPRGSTAIGMKETAATAAKFERQHMDSPDLGTDGDDDK 197

RESULT 19  
US-09-609-161B-22  
; Sequence 22, Application US/09609161B  
; Patent No. 6436682  
; GENERAL INFORMATION:  
; APPLICANT: Bryan, Bruce  
; APPLICANT: Szent-Gyorgyi, Christopher  
; APPLICANT: PROLUME, LTD.  
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE  
; TITLE OF INVENTION: AND FLUORESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS,  
; FILE REFERENCE: 24729-121B  
; CURRENT APPLICATION NUMBER: US/09/609.161B  
; CURRENT FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: 09/277,716  
; PRIOR FILING DATE: 1999-03-26  
; PRIOR APPLICATION NUMBER: 60/102,939  
; PRIOR FILING DATE: 1998-10-01  
; PRIOR APPLICATION NUMBER: 60/089,367  
; PRIOR FILING DATE: 1998-06-15  
; PRIOR APPLICATION NUMBER: 60/079,624  
; PRIOR FILING DATE: 1998-03-27  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 22  
; LENGTH: 382  
; TYPE: PRT  
; ORGANISM: Artificial Sequence: fusion protein  
; FEATURE:  
; OTHER INFORMATION: Cellulose Binding Domain (CBD)-Gaussia<sub>1</sub> luciferase fusion prote  
US-09-609-161B-22

```

; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/946.475
; FILING DATE: 08-OCT-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/027,973
; FILING DATE: 08-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50560
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
;
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 711 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
; US-08-946-475-9
;
; Query Match 6.6%; Score 151; DB 3; Length 711;
; Best Local Similarity 22.9%; Pred No. 0.0021;
; Matches 68; Conservative 39; Mismatches 80; Indels 110; Gap
;
; QY 2 HHHHHSSGLVPRGSGM-----KETAAAKFERQHMDSPLDGLDQDDKAMADIGSIE
; | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; DB 5 HHHHHSSGLVPRGSHMTADNLNVIVESPAKAKTTEKVLGK-----KYKVIASMGHVR
;
; QY 53 G--RGTMATSVTMDNLISGFENEDYIVILKPLAGVYRSILKKQIE--KNITFN-----1
; | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; DB 58 DLPRSQMGVD-TEDN----YEPY--ITIRGQPVVKELKHAKKAKNVFIASDPDRGE
;
; QY 102 -----LNLDILNSRLKKRKFKYFLDVLDSLMQFKHISSNEYIIEDSFKLLNSEQKN
; | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; DB 111 AIAWHLSKILEDSKENRVVNEITKDAVKESFNPREIEMN-----LVDAQQAR
;
; QY 153 TLKS-----YKYTKESV-----ENDI-----1
; | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; DB 162 RILDRLVGYNISPLVLRKKYKGLSAGRVQSVALRLVIDREINRFPKPEYWTIEGEF
;
; QY 170 ---KFAQEGISY-----YKVLAKYKDDLESIKKVKKEKPKFSSSPPTT 211
; | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; DB 222 KSKFNKFLHYKPKFKLTKTKDVEKITAALDQGFETVNTVTKKERTNPNPPTT 278
;
;
; RESULT 22
; US-09-340-479-9
; Sequence 9, Application US/09340479
; Patent No. 627A139
; GENERAL INFORMATION:
; APPLICANT: Gwynn, Michael
; APPLICANT: Kallender, Howard
; APPLICANT: Palmer, Leslie
; TITLE OF INVENTION: TOPOISOMERASE I
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA

```

```

; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA: US/09/340.479
; FILING DATE:
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/946.475
; FILING DATE: 08-OCT-1997
; APPLICATION NUMBER: 60/027,973
; FILING DATE: 08-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50560
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 711 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-340-479-9

Query Match 6.6%; Score 151; DB 4; Length 711;
Best Local Similarity 22.9%; Pred. No. 0.0021;
Matches 68; Conservative 39; Mismatches 80; Indels 110; Gaps 13;

Qy 2 HHHHHSSGLVPRGSM-----KETAARKFERQHMDSPDLGTDHDDDDKAMADIGSIE 52
Db 5 HHHHHSSGLVPRGSHMTLADNLVIVSPAKATIEKYLGR-----KYKVIASMGHVR 57

Qy 53 G--RGTMAISVTMDNILSGFENEYDVIYKPLAGVYRSLKKOIE--KNIFTN----- 101
Db 58 DLPRQMGVD--TEDN----YEPKY--ITIRGKGPVVKELKKHAKAKNVFLASDPDRGE 110

Qy 102 -----LNLNDILNSRLKKRYFLDVLSDLMQFKHISNEYIIEDSPKLLNSEQKN 152
Db 111 AIAWHLKILEDSKENRVVFEITKDAVKESFKNPREIEMN-----LVDAQOAR 161

Qy 153 TLLKS-----YKIKESV-----ENDI----- 169
Db 162 RILDLRLVGYNTSPVLWKVKKGLSAGRVOSVALRLVLDRENEIRNFKPEYWTIEGEPRY 221

Qy 170 ---KFAQEGISV-----YEVKLYAKYKDDLESIKKVIKEKEKFPSSPPTT 211
Db 222 KKSRENKFLHYKKNKPFKTKKDEKITAALDGDGDFEITNTVKKEKTRNANPFTT 278

RESULT 23
US-09-134-001C-5178
; Sequence 5178, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134.001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; NUMBER OF SEQ ID NOS: 5674

Qy 86 YRSLKKQIE---KNIFTNLMNDILNSRLKKRYFLDVLSDLMQFKHISNEYIIED 141
Db 6361 YNNALKQAEDLIINSSNPNLAQDITNA--LNNIKQAQNLHGAQKLDQDN--TTNQAI-- 6416

Qy 141 DSFKLLNSEQKNTLLKSYKIKESVENDIKFAQEGISYEVKLYAKYKDDLESIKKVIKEE 200
Db 6417 GNLNLPQKDALIQ-----AINGATSRDQVAEKLK--EAEALDAMKQL 6460

Qy 201 KEKFPSSPPTTPPSPAKTDEQKESKFLPFLTNITLNNLVN-----KIDDYLINLKA 254
Db 6461 EDQVNDQDISNSPFFINEDSDKQTKYNDKTAQAKKEIINQTSNPLDKQKLTADTLQNKD 6520
```

```

; SEQ ID NO 5178
; LENGTH: 1010
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5178

Query Match 6.1%; Score 139.5; DB 4; Length 1010;
Best Local Similarity 22.2%; Pred. No. 0.024;
Matches 71; Conservative 58; Mismatches 112; Indels 79; Gaps 14;

Qy 86 YRSLKKQIEKNIFTNLMNDI--LNSRLKKRYFLDVLSDLMQFKHISNEYIIED 141
Db 348 YHNEIKGFQKOLEHLSRENEITQFNYLEKNQVFFNOLD-----KLISSYQOKPVIIE 401

Qy 142 SFKLLNSEQKNTLLKSYKIKESVENDIKFAQEGISYEVKLYAKYKDDLESIKKVIKEE 201
Db 402 EIKRLYSEYNDLITKKEELTKEMNNKNKDFP-----IIEHYTEIYKLLKIIDE-- 450

Qy 202 EKFPSSPPTTPPSPAKTDEQKESKFLPFLTNITFLY-----NNLVNKIDDYILINLKA 254
Db 451 -----SERQKDEKLFQKLDKSSYLSKLKKEKQELNEIESTITNIDA 494

Qy 255 KIDCNVEKDEAHVKITKLSOLKAIDDKIDLFKNP-----YDPEAIKKLINDDTKKDM 307
Db 495 TLIDUNDKRD---FVNEIKSAMSIGDTCPCGNIHSLGEHIDFESTAQ-----KNNK 544

Qy 308 LGKLLSTGLVQNFNPNTIISKIEGKFQDM-----LNISQHCQVKKQCPENSGCFRHLDE 361
Db 545 IKRLSKKV--KIRDEIILK--IETRIEELHNRELNELFEKQE--KKDISELQKQLNLHLNQ 598

Qy 362 REECKLLNYKQEGDKCVEN 381
Db 599 -----LKDEQQSINKLVEN 612

RESULT 24
US-09-134-001C-3159
; Sequence 3159, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134.001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3159
; LENGTH: 10182
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3159

Query Match 5.8%; Score 133; DB 4; Length 10182;
Best Local Similarity 24.2%; Pred. No. 1.5;
Matches 68; Conservative 49; Mismatches 126; Indels 38; Gaps 10;

Qy 86 YRSLKKQIE---KNIFTNLMNDILNSRLKKRYFLDVLSDLMQFKHISNEYIIE 140
Db 6361 YNNALKQAEDLIINSSNPNLAQDITNA--LNNIKQAQNLHGAQKLDQDN--TTNQAI-- 6416

Qy 141 DSFKLLNSEQKNTLLKSYKIKESVENDIKFAQEGISYEVKLYAKYKDDLESIKKVIKEE 200
Db 6417 GNLNLPQKDALIQ-----AINGATSRDQVAEKLK--EAEALDAMKQL 6460

Qy 201 KEKFPSSPPTTPPSPAKTDEQKESKFLPFLTNITLNNLVN-----KIDDYLINLKA 254
Db 6461 EDQVNDQDISNSPFFINEDSDKQTKYNDKTAQAKKEIINQTSNPLDKQKLTADTLQNKD 6520
```





Db 81 GLISGNGTEAKVETN-----EKFKEIFGSGKDLKDLTDTGTFKGL 126  
Qy 108 LNSRLKRRKYFLDVLSDLMQFKHISNE-----YIIDS--FKLNSQKNTLLKSY 158  
Db 127 IESAAGK--LKDLLE--KFKLFEDETKAGRVKEILTDSNAKEILTNEVAKEVLKSD 182  
Qy 159 KYIESVENDIKFAQEGISYIEKV-----LAKYKDDLES-----IK 194  
Db 183 KF-KEAITGDKGDKALKEILTCDKFEKAVTNGKDKILGILTDSTGKFKELIESTSKDILK 241  
Qy 195 KVEEKEKPPSPPTPPSPAKTDEQKESKFLPFTNI-----ETLYNNLVNKIDD 247  
Db 242 EILDTNGFKGLLEST--GKEVKELLIDGKFKDLTDTATKAGYVKEILTNDTAKEV-- 297  
Qy 248 YLNLKAKINDCNVKEDEAHKTKLSLKAIDDKIDLFKNPYDEAI-----KKLNDND 302  
Db 298 -LTDQAK-----EVLKDSSTAKDILKDTNAAV-----LKNSTAKEILTNTQAKEVLTDG 346  
Qy 303 TKDMLGKLL-----STG-----LVQNFNPTIISKLI-----EGK 332  
Db 347 TSKEVLKEILTCDKFEKAVTNGDKDILGILTDSTGKFKELIESTGKDKILKDLTDTGK 406  
Qy 333 FQDMLNLSHQCVKKQCPENSGCGRFHLDE---REECRCLL 369  
Db 407 FKELIEVLVKNKLAKEILTDTNGFKGLVEGAGKDEAKAVL 446  
RESULT 28  
US-09-134-001C-4378  
; Sequence 4378, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 4378  
; LENGTH: 878  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4378

Query Match 5.6%; Score 126.5; DB 4; Length 878;  
Best Local Similarity 20.7%; Pred. No. 0.19;  
Matches 82; Conservative 57; Mismatches 134; Indels 123; Gaps 16;

Qy 15 GSGMKETAAAKF--ERQHMSPDLGTDDDKAMADIGSIEGRGTMAISVTMDNLSGFEN 72  
Db 458 GIGKLEVOGKISREEQYQNTKQIIEDN-----ISIT--NIIIPNE 499  
Qy 73 EYDVYLKPLAGVYRSLLKQIEK-NITFNLNLNLSRLKRRKYFLDVLSDLMQFKH 131  
Db 500 NYN-----QQIKHVEIFDSIKNTDHNKKLKE----- 527  
Qy 132 ISSNEYIIDSFKLLNSEQNTLLSKYIKESVENDIKFAQEGISYIEKVLAKYKDDLE 191  
Db 528 -----IKSMFTDLKDTTONEIEKIYNVWKEKKKTEKINRAIKSLDDIEGKTRED-- 578  
Qy 192 SIKKVIKEEKFPSPPTPPSPAKTDEQKESKFLPFTNIETLYN-----NLVKN 244  
Db 579 -IAHEYETQKQITSIEPL-----ETQLSRVKSITETLENRIQLKEDLKEI 624  
Qy 245 IDDYLYNLK---AKINDCNVKEDEAHKTKLSLKAIDDKIDLFKNPYDEAI-----DKIDLKFN-- 288  
Db 625 FDEQLKLNRCVKKINNRLLKK-QVNIKIQPYANVNNLIEFLKEENGLDGDLTKWIKNHQ 683

Qy 289 PYDEAIAKKLINDTKKMDLGLKLLSTGLVQNFNPTIISKL-----IEGKFQDML 337  
Db 684 SFNPFKFIKLKDRDSEAIYBEYKDSGLKKHTADILSNMLYERILKLESIELENIIDRL 743  
Qy 338 NISOHQCVKKQCPENSGCGR---HLDRECKCLLN 370  
Db 744 NVGSKDTK-----FRSLNHLKSGQOCTAILN 770

## RESULT 29

US-08-755-587-186  
; Sequence 186, Application US/08755587  
; Patent No. 6045997  
; GENERAL INFORMATION:  
; APPLICANT: Futreal, Phillip A  
; APPLICANT: Wooster, Richard F  
; APPLICANT: Ashworth, Alan  
; APPLICANT: Stratton, Michael R  
; TITLE OF INVENTION: Materials and methods relating to the  
; TITLE OF INVENTION: Identification and sequencing of the BRCA2 cancer  
; TITLE OF INVENTION: susceptibility gene and uses thereof.  
; NUMBER OF SEQUENCES: 222  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Bell Seltzer Park & Gibson  
; STREET: 310 OCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107  
; CITY: Raleigh  
; STATE: NC  
; COUNTRY: USA  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/755,587  
; FILING DATE: 25-NOV-1996  
; PRIOR APPLICATION DATA: GB 9523959.6  
; APPLICATION NUMBER: GB 9523959.6  
; FILING DATE: 23-NOV-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9525555.0  
; FILING DATE: 14-DEC-1995  
; PRIOR APPLICATION DATA: GB 9617961.9  
; APPLICATION NUMBER: GB 9617961.9  
; FILING DATE: 28-AUG-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kenneth D Sibley  
; REGISTRATION NUMBER: 31,665  
; REFERENCE/DOCKET NUMBER: 5405-135  
; INFORMATION FOR SEQ ID NO: 186:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1494 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
US-08-755-587-186

Query Match 5.6%; Score 126.5; DB 3; Length 1494;  
Best Local Similarity 21.7%; Pred. No. 0.38;  
Matches 94; Conservative 75; Mismatches 188; Indels 77; Gaps 20;

Qy 37 GTDDDDKAMADIGSIEGRGTMAISVTMDNLSGFEN-----EYDVYLKPLAGVYRSLLK 90  
Db 736 GNTQIKEGLSDLTCLVMKABETSHTVMTSNKQOLTANTGQIKDFDTFYL-----SFQ 787  
Qy 91 KQTEKNFTNENLN-----DILNSRLKRRKY--FLDVLSDLMQFKHIS-----SNEYIED 141  
Db 788 TASKNIRVRESLNKARSLLNQKWTBEELNFNFSLSNSELPGIDIKKTDISNHEVIE- 846  
Qy 142 SFKLLNSEQNTLLSKYIKESV-----ENDIKFAQEG--ISYIE---KVLAKYKD 188  
Db 847 -----NTERKDKITKESLIGTENILLIQORPESKIKKIKRESAVLGFTASGKKIEITKE 901  
Qy 189 DLESIKKVIKEEKFPSPPTPPSPAKTDEQKESK-----FLPFLNIETLY 238

```

Db 902 SLDKVNLF--EKEQDNSEITNFSHRGAKMSKORECKDGRACGTTEITTTPEYEETH 960
QY 239 NNL-VNKIDYLNKAKINDCNVDEAHVKITKLSDLKAIDDKIDLFKNPYDFAEAIKK 297
Db 961 SSLEKKVSNELAAALRRLSLNLYKOTENKLSIDHA-----SOKVDVHENT-EKETAK- 1013
QY 298 LINDDTKQMLGKLLSTGLVQNPNTIISKLEGFQDMLNISQHCVRKQCPENSGCPR 357
Db 1014 -----KPTMYTNOSTYSIAENSPLT-FQDTEKEF-----SVSEASLFEAKKWLREGEWD 1062
QY 358 HLDERECKLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSRKKITCECTK 417
Db 1063 DQSERINA-AKVCKEYPDDYVENPNCNCGNSAITSNENKHLSEKOGST---YLSNSTM 1118
QY 418 PDSYPLFDGIFCSS 431
Db 1119 SNSYSYHPC-FCHS 1131

RESULT 30
US-08-480-640A-115
: Sequence 115, Application US/08480640A
: Patent No. 6033904
: GENERAL INFORMATION:
: APPLICANT: Cochran, Mark D.
: APPLICANT: Junker, David E.
: TITLE OF INVENTION: Recombinant Swinepox Virus
: NUMBER OF SEQUENCES: 225
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: John P. White
: STREET: 1185 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: USA
: ZIP: 10036
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/480,640A
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: White, John P
: REGISTRATION NUMBER: 28,678
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 278-0400
: TELEFAX: (212) 391-0525
: INFORMATION FOR SEQ ID NO: 115:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 677 amino acids
: TYPE: amino acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Swinepox virus
: STRAIN: Kasza
: INDIVIDUAL ISOLATE: S-SPV-001
: IMMEDIATE SOURCE:
: CLONE: 515-85.1
: POSITION IN GENOME:
: MAP POSITION: -23.2
: UNITS: %G
US-08-480-640A-115

Query Match 5.5%; Score 126; DB 3; Length 677;
Best Local Similarity 22.8%; Pred. No. 0.15;
Matches 78; Conservative 65; Mismatches 135; Indels 64; Gaps 19;

```

```

Matches 78; Conservative 65; Mismatches 135; Indels 64; Gaps 19;
QY 75 DVYILKPLAGVYRSLLKKQIEKNIFTF-----NLNLDILNSRLKRRKYFLDVL---SDLM 127
Db 185 DAFHIPPTHTSLRSL-RDINRIELKKYPNNIIDYISDIKSNSSFIHLHMIISNMF 243
QY 128 OFKHISSENYI--IEDSFKLLNSEQ-KNTLLKSYK-----IKESVENDIKFAQEGISYVE 180
Db 244 PAIIPSVNDFISTVWDKRLINMYGKCVAMFSYDINIMDLESDDSDYIFIEKNIYD 303
QY 181 KVLAKYKDDLESITKVKIEKEKFPSPPTTPSPAKTDEQKESKFLPFLTNIETLYNN 240
Db 304 ---VKRDFANMIRDKVREKNRILFT-----KCEDIIRYIKLFS-----KNRIND 346
QY 241 LVNKIDYLNKAKINDCNVDEAHVKITKLSDLKAIDDKIDLFKNPYDFAEAIKKLIN 300
Db 347 ENNKVEVLHID-----NVSKNKK-----LSLSDISLMDQFRL--NPCTIRNI--LLS 392
QY 301 DDTKKOMLGLKLLSTGLVQNF-----PNTIISKLEIEKGF-QDMLNISQHCVRK--KQCPEN 352
Db 393 SATIK---SKLLALRAVKNNKCYSLTNVSMVKKIKGVIMDMVDYISTNLIKHYHKQLYDK 449
QY 353 SGCFRHLDERECKC-----LLNYKQEGDKCVENPNPTCNE 388
Db 450 MSTFEYKARDIKSKCSCSDSITHIHYETTSCI-NYKSTDND 490

RESULT 31
US-08-480-640A-193
: Sequence 193, Application US/08480640A
: Patent No. 6033904
: GENERAL INFORMATION:
: APPLICANT: Cochran, Mark D.
: APPLICANT: Junker, David E.
: TITLE OF INVENTION: Recombinant Swinepox Virus
: NUMBER OF SEQUENCES: 225
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: John P. White
: STREET: 1185 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: USA
: ZIP: 10036
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/480,640A
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: White, John P
: REGISTRATION NUMBER: 28,678
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 278-0400
: TELEFAX: (212) 391-0525
: INFORMATION FOR SEQ ID NO: 193:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 677 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-480-640A-193

Query Match 5.5%; Score 126; DB 3; Length 677;
Best Local Similarity 22.8%; Pred. No. 0.15;
Matches 78; Conservative 65; Mismatches 135; Indels 64; Gaps 19;

```

```

QY 75 DVYILKPLAGVYRSLLKKQIEKNIFTF-----NLNLDILNSRLKRRKYFLDVL---SDLM 127
Db 185 DAFHIPPTHTSLRSL-RDINRIELKKYPNNIIDYISDIKSNSSFIHLHMIISNMF 243

Query Match 5.5%; Score 126; DB 3; Length 677;
Best Local Similarity 22.8%; Pred. No. 0.15;
Matches 78; Conservative 65; Mismatches 135; Indels 64; Gaps 19;

```

Qy	128	QFKHISSNEVI--LEDSFKLLNSEQ-KNTLLKSKYK-----IKESVENDIKFAQEGISYSE	180
Db	244	PAIIPSVNDFISVVDKRLINMYGKCVAMFSTOINWDLSDSDSIVIFTEKNISYD	303
Qy	181	KVIAKYKDDLESKKVKEKEKPPSPPTTSPSPAKTDEQKKESKFLPFLTNIETLYNN	240
Db	304	--VKCRDFANMIRDVKREKNRLT-----KCEIIRYIKLFS-----KNRIND	346
Qy	241	LVNKIDDYLNLKAKINDCNDVKEDEAHVKITKLSDLKAJDDKIDLFKNPYDFAEIKKLIN	300
Db	347	ENNKVEEVLHID-----NVSKNNK-----LSLSDSLSSLDQFRL--NPTCTIRNI--LLS	392
Qy	301	DDTKKMDLGMGLKLLSTGLVQNF-----PNTIISKLEIGKF-QDMLNISTOHOCVK--KQCPEN	352
Db	393	SATIK-----SKLLAURALVKNCKVSLTNVSMYKIKGVIMDVTISTNILKYHKLQYDK	449
Qy	353	SGCFRHLDERECKC-----LLNYKQEGDKVCENPNPTCNE	388
Db	450	MSTFEYKRDIKSKSCICSDSIPIHHIYETTSCI-NYKSTDN	490

```

RESULT 32
US-08-295-802-1115
; Sequence 115, Application US/08295802
; Patent No. 6127183
; GENERAL INFORMATION:
; APPLICANT: Cochran Ph.D., Mark D
; APPLICANT: Junker M.S., David E
; TITLE OF INVENTION: Recombinant Swinepox Virus
; NUMBER OF SEQUENCES: 188
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; 7TD: 10112

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

US-08-295-802-115

```

Query Match          5.5%; Score 126; DB 3; Length 677;
Best Local Similarity 22.8%; Pred. No. 0.15;
Matches 78; Conservative 65; Mismatches 135; Indels 64; Gaps 19;

Qy 75 DVIYKPLAGVYRSLKKOIEKNITFF-----NLNLDILNRLKKRKVFLDVLV---SDLM 127
      ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 185 DAFPHITHTLSRL-RDINRIETLLKKYPNNIIDVYSIKSSSFIIHLMIISNMF 243
      ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Qy 128 QFKHISSEVY-IEDSFKLLNSQ-KNTLLKSYY-----IKESVENDIKFAOGISYYE 180
      ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 244 PAIIPSVNDFISTVVDKRLINMYGIKCVAMFSYDINIMDLESDDSDYIFIEKNISYD 303
      ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Qy 181 KVLAKYVDDLESIKKVIKEKEKPPSPPTTPPSPAKTEQKKESKFLPFLTNIETLYNN 240
      ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 304 ---VKCRDFANIRDKVYREKNRILTT-----KCEDLIIRVIKLF-----KNRIND 346
      ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Qy 241 LVNKIDDYLLNKAINDCNVEKDEAHVKITKLSDLKAIDDKIDLFPKNPYDFEAIKKLIN 300
      ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 347 ENNKVEEVLTHID-----NVSKNKK-----LSLSDISSLMDQFRL--NPCTIRNI--LLS 392
      ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Qy 301 DDTKKMDLGLKLLTGLVONF-----PNTIISKIEGKF-QDMLNISOHQCVK--KQCPEN 352
      ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 393 SATIK-----SKLLALRAVKNKCYSLTNVSYKKIKGVIMDVMDYDYSITNLKLYHKQLYDK 449
      ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Qy 353 SGCFRHLDEREECK-----LLNYKQEGDKCVENPNPTCNE 388
      ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 450 MSTFEYKRDIKSKCSICSDSIITHIHYETTSCI-NYAKSTDND 490
      ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

RESULT 33
US-08-686-968C-58
; Sequence 58, Application US/08686968C
; Patent No. 6221361
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D.
; APPLICANT: Junker, David E.
; TITLE OF INVENTION: Recombinant Swinepox Virus
; FILE REFERENCE: 39119-H/JML
; CURRENT APPLICATION NUMBER: US/08/686,968C
; CURRENT FILING DATE: 1996-07-25
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 58
; LENGTH: 677
; TYPE: PRT
; ORGANISM: Swinepox virus
US-08-686-968C-58

```

```

Query Match      5.5%; Score 126; DB 4; Length 677;
Best Local Similarity 22.8%; Pred. No. 0.15;
Matches 78; Conservative 65; Mismatches 135; Indels 64; Gaps 19;

QY 175 DVLYKPLAGVYRSLKKQIEKNITFF-----NLNLNDILNRLSKRKYFLDVLV---SDLM 127
      | : : | | : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 185 DARHIPPTHTLSRL-RDINRIITELLKKYPNNNIIDYISDIKSNSSFIIRHLAMITSNMF 243
      | : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 128 QFKHISSENEYI--IEDSEFKLLNSEQ-KNTLLSKYK----IKESVENDIKFAOEGISYE 180
      | : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 244 PAIIPSWNDIFSTVVDKDRLINMYGICVAMFYSDINIMIDLESLLDSDSYFIEKNTSIYD 303
      | : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 181 KVLAKYKDDLESIAKKVIEKEKFPSPPTTPPSPAKTDBQKESKFLPFLTNIETLYNN 240
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 304 ---VKRDFANMIRDKVKREKNRILTT-----KCEDIIRYIKLFS-----KNRIND 346
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 241 LVNKIDDYLYNLRAKINDCNVEXDEAHVKTITKLSDLKAIDDKIDLKNPYDFFAIFKLIN 300
      | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 347 ENNKVEEVLIHID-----NVSKNNK-----LSLSDISLLMDQFRL--NPCTIRNI--LLS 392
      | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
QY 301 DDTFKOMLGLKLLSTGLVONE-----PNTIISKIEGKF-QOMLNISOHCVK--KQCPEN 352
      | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 393 SATIK-----SKLLALRAVKNWKCYSLTNVSMYKKIKGVIVMDVDYISTNLTKYHKQLYDK 449
      | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
QY 353 SGGFRHLDEREECK-----LLNYKQEGDKVCNPNPNTCNE 388

```



COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/488,237A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P  
REGISTRATION NUMBER: 28,678  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525  
INFORMATION FOR SEQ ID NO: 193:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 677 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-488-237A-193

Query Match 5.5%; Score 126; DB 4; Length 677;  
Best Local Similarity 22.8%; Pred. No. 0.15;  
Matches 78; Conservative 65; Mismatches 135; Indels 64; Gaps 19;  
QY 75 DVILYKPLAGVYRSKQKQIEKNIFTF-----NLNLDILNSRLKKRKYFLDVL-----SDLM 127  
DB 185 DAFHIPPHTHSRL-RDINRIELKKYPNNIIDIYSDSIKSNSSFIHLHMIISNMF 243  
QY 128 QFKHSSNEYI--IEDSKLLNSEQ-KNTLLKSYKY----IKESVENDIKFAQEGISYVE 180  
DB 244 PAIIPSVNDFSTVVYDKRLINMYGKCVAMPSYDINMIDLESDDSDYIFIEKNISYD 303  
QY 181 KVLAKYKDDLESIKKVIKEKEKFPSSPTTPPSAKTDEQKESKFLPFLTNITLYNN 240  
DB 304 ---VKCRDFANMIRDKVREKNRIIT-----KCEDIIRYIKLFS-----KNRIND 346  
QY 353 SCGFRHLDERECKC-----LLNKKQEGDKCVENPNTCNE 388  
DB 450 MSTFEYKRDIKSKCKSCISDSITHIYETTSCI-NYKSTDND 490

RESULT 37  
US-08-375-992A-115  
Sequence 115, Application US/08375992A  
Patent No. 6328975  
GENERAL INFORMATION:  
APPLICANT: Cochran, Mark D.  
APPLICANT: Junker, David E.  
TITLE OF INVENTION: Recombinant Swinepox Virus  
NUMBER OF SEQUENCES: 220  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: John P. White  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/375,992A  
FILING DATE: Herewith  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P  
REGISTRATION NUMBER: 28,678  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525  
INFORMATION FOR SEQ ID NO: 115:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 677 amino acids  
TYPE: amino acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Swinepox virus  
STRAIN: Kasza  
INDIVIDUAL ISOLATE: S-SPV-001  
IMMEDIATE SOURCE:  
CLONE: 515-85.1  
POSITION IN GENOME:  
MAP POSITION: -23.2  
UNITS: %G  
US-08-375-992A-115

Query Match 5.5%; Score 126; DB 4; Length 677;  
Best Local Similarity 22.8%; Pred. No. 0.15;  
Matches 78; Conservative 65; Mismatches 135; Indels 64; Gaps 19;  
QY 75 DVILYKPLAGVYRSKQKQIEKNIFTF-----NLNLDILNSRLKKRKYFLDVL-----SDLM 127  
DB 185 DAFHIPPHTHSRL-RDINRIELKKYPNNIIDIYSDSIKSNSSFIHLHMIISNMF 243  
QY 128 QFKHSSNEYI--IEDSKLLNSEQ-KNTLLKSYKY----IKESVENDIKFAQEGISYVE 180  
DB 244 PAIIPSVNDFSTVVYDKRLINMYGKCVAMPSYDINMIDLESDDSDYIFIEKNISYD 303  
QY 181 KVLAKYKDDLESIKKVIKEKEKFPSSPTTPPSAKTDEQKESKFLPFLTNITLYNN 240  
DB 304 ---VKCRDFANMIRDKVREKNRIIT-----KCEDIIRYIKLFS-----KNRIND 346  
QY 241 LVNKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLKALDKIDLFKNPYDEATKLLIN 300  
DB 347 ENNKVEEVLHID-----NVSNNK-----LSLSDISSLMQDFRL--NPCTIRNI--LLS 392  
QY 301 DDTKKDMLGKLLSTGLVONF-----PNTIISKLEKGF-QDMLNISQHCQVK--KQCPEN 352  
DB 393 SATIK---SKLLALRAVNWKCYSLTNVSMYKKIKGVIVMDVYISTNILKYHKQLYDK 449  
QY 353 SCGFRHLDERECKC-----LLNKKQEGDKCVENPNTCNE 388  
DB 450 MSTFEYKRDIKSKCKSCISDSITHIYETTSCI-NYKSTDND 490

RESULT 38  
US-08-375-992A-193  
Sequence 193, Application US/08375992A  
Patent No. 6328975  
GENERAL INFORMATION:  
APPLICANT: Cochran, Mark D.  
APPLICANT: Junker, David E.  
TITLE OF INVENTION: Recombinant Swinepox Virus  
NUMBER OF SEQUENCES: 220  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: John P. White  
STREET: 1185 Avenue of the Americas

CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
FILING DATE: August 9, 1994  
APPLICATION NUMBER: US/08/375,992A  
FILING DATE: Herewith  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P  
REGISTRATION NUMBER: 28,678  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525  
INFORMATION FOR SEQ ID NO: 193:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 677 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-375-992A-193

Query Match 5.5%; Score 126; DB 4; Length 677;  
Best Local Similarity 22.8%; Pred. No. 0.15;  
Matches 78; Conservative 65; Mismatches 135; Indels 64; Gaps 19;  
QY 75 DVIYKPLAGVYSLKQTEKNIETP-----NMLNDILNSRLKRYFDVLE---SDLM 127  
DB 185 DAFHIPPTISLRSR-LRINRIELLLKYPNNIIDIYISDKSNSSFIHLHNIISNMF 243  
QY 128 QKHISSEYI--IEDSFLLNSEQ-KNTLLSKYK-----IKRESVNDIKFAEGISYE 180  
DB 244 PAIPSVNDFISTVDKDLINMGYKCVAMFSYDINMIDLESDDSDYIFIEKNISYD 303  
QY 181 KVLAKYKDDLESIKKVIKEKFKPSPPTTPSPAKTOEQKESKFLPLTNIETLYNN 240  
DB 304 ---VKCRDFANMRDVKREKNRLIT-----KCDIIRYIKLFS-----KNRIND 346  
QY 241 LVNKIDYILINKAKINDCNVEKDEAHVKITLSOLKAIDDKIDLKPNVDFEAIKKLIN 300  
DB 347 ENNKVEVLHID-----NVSKNNK-----LSLSDISSLMQFRL--NPCTIKNI--LLS 392  
QY 301 DDTKKDMLKLLSTGLVONF-----PNTIISKLEGGF-QDMLNISQHCVK--KQCPEN 352  
DB 393 SATIK---SKLLALRAVKNKCYSLTNVSMYKKIKGVIVMDVYISTNLLKYHKOLYDK 449  
QY 353 SGCFRHLDERECK-----LLNYKQEGDKCVENPNPTCNE 388  
DB 450 MSTFEYKDKIKSKCSCICSISITHIYETTSICI-NYKSTDND 490

RESULT 39  
US-08-287-959-1  
Sequence 1, Application US/08287959  
Patent No. 5639651  
GENERAL INFORMATION:  
APPLICANT: Weissbach, Lawrence  
APPLICANT: Bernards, Andre  
APPLICANT: Settlemann, Jeffrey  
TITLE OF INVENTION: GAP-RELATED GENE  
NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: U.S.A.  
ZIP: 02110

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
FILING DATE: August 9, 1994  
APPLICATION NUMBER: US/08/287,959  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Clark, Paul C.  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 00786/181001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1657 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-287-959-1

Query Match 5.5%; Score 125.5; DB 1; Length 1657;  
Best Local Similarity 17.5%; Pred. No. 0.52;  
Matches 69; Conservative 81; Mismatches 106; Indels 139; Gaps 16;  
QY 75 DVIYKPLAGVYSLKQTEKNIETPNNLNDILNSRL-----KKRYFDVLESD----- 125  
DB 808 EVYKIQSLARMHQA-RKRYRDRLOQYFRDHINDIIOAFIRANKARDYKTLINAEPPM 866  
QY 126 --LMQFKHI--SSEYIIEDSFKLLNSEQNTLLKSKYKIKESVENDIKF----- 171  
DB 867 VVYKRVHLLDQSDQDFQELDLKMKREEVITLIRS-----NQLENDLNMIDKIGLLVK 922  
QY 172 ----AQEGISYEVKYLAKYKDDLE-----SIKKVIKEKEKFPSPSP-----PTT 211  
DB 923 NKITLDQVYSHSKLTAKKNEQLSDMMINKQGGKALKSKREKLEAYOHLFYLLQTN 982  
QY 212 PPSPAKTDEQKESKFLPLTN-IEIYNNLVNKIDYLI-----NLKAKINDCN 260  
DB 983 PTYLAKLIFQMPONKSTKFMDSVIFTLYNYSNQREYLLRLFKTALOEIKSKVDQIQ 1042  
QY 261 --VEKDEAHVK-----ITKLSDLKAID--DKIDLEKN----- 288  
DB 1043 EIVTGNPTVIKVVSNRGARGONALRQILAPVVKEMDDKSLNIKTDPVDIYKSVNQM 1102  
QY 289 -----PYDFAIKKLLINDDTKKDM----- 307  
DB 1103 ESQTGEASKLPYDVTPEQALAHAEVKTRLDSSIRNMRAVTDKLSAIVSSVDKIPYGMRF 1162  
QY 308 LKLLSTGLVQNEPNT-----IISKLEGGKFD 335  
DB 1163 IAKVLKDSLHREKPPDAGEDELLKILGNLLYYRYMN 1197

RESULT 40  
US-08-235-836C-74  
Sequence 74, Application US/08235836C  
Patent No. 6248562  
GENERAL INFORMATION:  
APPLICANT: Dunn, John J.  
APPLICANT: Luft, Benjamin J.  
TITLE OF INVENTION: No. 6248562el Chimeric Proteins Comprising  
NUMBER OF SEQUENCES: 144  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Brookhaven National Laboratory  
STREET: Upton  
CITY: Upton  
STATE: NY

;; COUNTRY: USA  
;; ZIP: 11973  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patent in Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/235,836C  
;; FILING DATE: 29-APR-1994  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/148,191  
;; FILING DATE: 01-11-93  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Bogosian, Margaret C.  
;; REGISTRATION NUMBER: 25,324  
;; REFERENCE/DOCKET NUMBER: BNL93-28A  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (516) 282-7338  
;; TELEFAX: (516) 282-3729  
;; INFORMATION FOR SEQ ID NO: 74:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 700 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-08-235-836C-74

Query Match 5.5%; Score 125; DB 4; Length 700;  
Best Local Similarity 20.4%; Pred. No. 0.18; Indels 114; Gaps 16;  
Matches 84; Conservative 61; Mismatches 153;  
Qy 39 DDDKAMADIGSIEGRGTMAISVTMDNITLSCFENEYDVYIKPLAGVYRSIKKQIEKNIF 98  
Db 88 DQDKKASVDIFSIGSKSELDLSILNRLITG-----YLMKSFYERSAELIAKAIT 139  
Qy 99 TFNMLNDILNSRLKRYFLDVLDSLMQPKHISSENYIETDSFKLNSQKNTLLKSY 158  
Db 140 IYNA-----VYRGDLDYK-----EFYIEASLKSITKENAG--LSRVY 175  
Qy 159 -----KYTKESVENDIKFAQEGISYVERVLAKYKDDLESI--KKVI---- 197  
Db 176 SOWAGKTOIFIPLKNNILSGVESDI-----DIDSLVTDKVVAAALL 216  
Qy 198 --KEEKEFPSPPTTSPAKTDEQKESKFLPPLNIETLYNNLVNKIDDYILINKAK 255  
Db 217 SENESGVNFARDITDIQETHKAQDKIDIE---LDNIHESDSNITETIENLRDLEKA 272  
Qy 256 INDCNVEKDEAHVKITKLS----DLKAIT-----DKIDLKPNPYDF-----EAIKKLIN 300  
Db 273 TDEEHKKEIESQVDAKKKQKEELDKAIDLDKAQOKLDFDAEDNLDIQRDVTVREKLQENIN 332  
Qy 301 DDTKKDMLGKLLSTGLVGNFPNTIISKLEGKFO--DMLNISQHCYKQKQCPENSGCFRH 358  
Db 333 ETNKEKNLPK-----PGDVSSPKVDKQLQIKESLEDLQEQ--LKETGDEN----- 375  
Qy 359 LDERECKCLLNYKQEGDKVCVENPNTCTNENNGCGADATCTEEDSGSSRRK 410  
Db 376 --QKREIEKQIEIKKDEKLLKSKDDRKASKDQKALDLDRELSKASKEKSK 425

Search completed: May 12, 2003, 10:20:56  
Job time : 43 secs



GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 19, 2003, 15:54:27 ; Search time 131 Seconds  
(without alignments)  
4243.995 Million cell updates/sec

Title: US-10-057-531A-2

Perfect score: 2275

Sequence: 1 MHHHHHHHSLVPRSGMKE.....TCECTKPSYLFDFGIFCSS 431

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 810007 seqs, 644969091 residues

Total number of hits satisfying chosen parameters: 1620014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cgn2.1/USPTO\_spool/US10057531/runat.12052003.091033.21772/app.query.fasta.1.583  
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=40 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US10057531@cgn.1.1.96 -runat.12052003.091033.21772  
-NCPU=6 -ICPU=3 -NO\_XLPHY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications\_NA:

1: /cgn2.6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2.6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2.6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2.6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2.6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2.6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2.6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2.6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2.6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2.6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2.6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
12: /cgn2.6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2.6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
14: /cgn2.6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	Length	ID	Description
1	1928	84.7	1235	9 US-10-098-514-13	Sequence 13, Appl
2	1911	84.0	1149	9 US-10-098-514-3	Sequence 3, Appli
3	1900	83.5	1149	9 US-10-098-514-1	Sequence 1, Appli
4	1175	51.6	5917	9 US-10-087-464-9	Sequence 9, Appli

5	1165.5	51.2	1131	9 US-10-087-464-58	Sequence 58, Appl
6	1137	51.2	1137	9 US-10-087-464-54	Sequence 54, Appl
7	1161.5	51.1	1065	12 US-10-082-018-1	Sequence 1, Appli
8	1161.5	51.1	1088	12 US-10-082-018-2	Sequence 2, Appli
9	1154.5	50.7	1142	12 US-10-082-018-8	Sequence 8, Appli
10	527	23.2	330	10 US-09-134-333-9	Sequence 9, Appli
11	527	23.2	342	10 US-09-134-333-6	Sequence 6, Appli
12	527	23.2	354	10 US-09-134-333-4	Sequence 4, Appli
13	523	23.0	279	10 US-09-134-333-3	Sequence 3, Appli
14	523	23.0	291	10 US-09-134-333-1	Sequence 1, Appli
15	523	23.0	387	10 US-09-134-333-7	Sequence 7, Appli
16	518	22.8	343	9 US-10-087-464-59	Sequence 59, Appl
17	306.5	13.5	1551	10 US-09-804-626-3	Sequence 3, Appli
18	304.5	13.4	1557	10 US-09-804-626-1	Sequence 1, Appli
19	278.5	12.2	2472	9 US-10-046-583A-6	Sequence 6, Appli
20	262.5	11.5	1065	9 US-10-012-896-1010	Sequence 1010, Ap
21	168	7.4	150	9 US-10-098-514-15	Sequence 15, Appl
22	144	6.3	2988	10 US-09-815-242-4196	Sequence 4196, Ap
23	144	6.3	3030	10 US-09-815-242-8043	Sequence 8043, Ap
24	140.5	6.2	2943	10 US-09-888-615-6	Sequence 6, Appli
25	137	6.0	3057	10 US-09-974-300-2660	Sequence 2660, Ap
26	132	5.8	2760	9 US-09-893-519A-146	Sequence 146, App
27	131.5	5.8	2847	9 US-10-267-311-20	Sequence 20, Appl
28	131	5.8	2853	10 US-09-764-864-373	Sequence 373, App
29	129	5.7	2341	10 US-09-881-752A-147	Sequence 147, App
30	129	5.7	6386	9 US-10-098-841-40	Sequence 40, Appl
31	128	5.6	1176	9 US-10-046-935-2238	Sequence 2238, Ap
32	128	5.6	1176	9 US-10-146-502-2238	Sequence 2238, Ap
33	127	5.6	4198	10 US-09-964-824A-263	Sequence 263, App
34	127	5.6	4967	12 US-10-044-090-102	Sequence 102, App
35	126.5	5.6	6773	10 US-09-864-864-336	Sequence 336, App
36	126	5.5	1678	10 US-09-974-300-2643	Sequence 2643, Ap
37	125.5	5.5	7573	10 US-09-880-107-2195	Sequence 2195, Ap
38	125	5.5	142	9 US-10-098-514-16	Sequence 16, Appl
39	123	5.4	2405	9 US-09-298-523B-70	Sequence 70, Appl
40	123	5.4	3463	9 US-09-298-523B-59	Sequence 59, Appl
41	122.5	5.4	8133	9 US-10-011-366-5	Sequence 5, Appli
42	121.5	5.3	12313	9 US-10-171-311-7	Sequence 7, Appli
43	121.5	5.3	12337	9 US-10-171-311-5	Sequence 5, Appli
44	121.5	5.3	12438	9 US-10-171-311-3	Sequence 3, Appli
45	121.5	5.3	12462	9 US-10-171-311-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1  
US-10-098-514-13  
; Sequence 13, Application US/10098514  
; Publication No. US20020194648A1  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Sandra P  
; APPLICANT: Christopher, David A  
; APPLICANT: Vine, Benjamin  
; APPLICANT: Su, Wei-Wen  
; APPLICANT: Bugos, Robert  
; TITLE OF INVENTION: PLASMIDIUM FALCIPARUM MEROZOITE SURFACE PROTEIN-1 MALARIA PROFI  
; FILE REFERENCE: A-71339/RET/TAL/NBC  
; CURRENT APPLICATION NUMBER: US/10/098,514  
; CURRENT FILING DATE: 2002-08-06  
; PRIOR APPLICATION NUMBER: US 09/500,376  
; PRIOR FILING DATE: 2000-02-08  
; PRIOR APPLICATION NUMBER: US 60/274,599  
; PRIOR FILING DATE: 2001-03-09  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 13  
; LENGTH: 1235  
; TYPE: DNA  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: synthetic  
US-10-098-514-13

Alignment Scores:

Pred. No.:	1.9e-169	Length:	1335
Score:	1928.00	Matches:	368
Percent Similarity:	98.67%	Conservative:	2
Best Local Similarity:	98.13%	Mismatches:	5
Query Match:	84.75%	Indels:	0
DB:	9	Gaps:	0

US-10-057-531A-2 (1-431) x US-10-098-514-13 (1-1235)

Qy	56	ThrMetAlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAsp	75
Db	73	ACAGACGCGATCTCTGTACTATTGGACAACATCTCTCAGTGGCTTCGAGAACGAGTACGAC	132
Qy	76	ValIleTyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLys	95
Db	133	GTAACTACCTAAAGCCCTTGCGGTGTCTACCGTTTCATTGAGAGAACAGATAGAAAG	192
Qy	96	AsnIlePheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArg	115
Db	193	AATATATTTCAGTTCACCTTCAACCTTAAATAGACATCCTCAACTCGCGCTCAAGAAGCGA	252
Qy	116	LysTyrPheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsn	135
Db	253	AAATACTTCTCGACGTGTGGAAATCCGACCTTATGCAATTTAAGCACATTTAGTCTTAAC	312
Qy	136	GluTyrIleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeu	155
Db	313	GAGTACATCATAGGACAGCTTCAAGCTCTTGAATTCAGACAGAGAACACCCTCCTA	372
Qy	156	LysSerTyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGly	175
Db	373	AACTCTCTACAAATACATTAGGAGTCTGTGTAGAACGACATCAAGTTCCGCCAGGAAGA	432
Qy	176	IleSerTyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLys	195
Db	433	ATTAGCTACTATCAGAAAGTCTCTGGCTTAATACNAGACGACCTTGGAAAGCATTTAAGAG	492
Qy	196	ValIleLysGluLysGluLysPheProSerSerProThrThrProProSerPro	215
Db	493	GTAATCAAGAAGAGAGAGAAAGTTTCCGAGCTCTCCACCCACAACTCCCCCATCGCCT	552
Qy	216	AlaLysThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGlu	235
Db	553	GCAAGACCCGACGAGCAGAAAAAGAAAGTAAGTCTCTTCATCTCCACCAACATCGAA	612
Qy	236	ThrLeuTyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLys	255
Db	613	ACTCTATATAACAACCTGGTGAAAGATTGATGACTTAACTAACTTGAAGCGCAAA	672
Qy	256	IleAsnAspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAsp	275
Db	673	ATTAATGACTGTAACTCGAAAAAGGATGAAGCCCGCTTAAGATCACCAAGCTTTCCGAT	732
Qy	276	LeuLysAlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIle	295
Db	733	CTCAAGCCATCAGCGNTAGATTGACCTGTTTAAGAACCCACACGATTTCCAGCGCAATC	792
Qy	296	LysLysLeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGly	315
Db	793	AAAAAGTTGATCAACGACGATACTAAGAAAGACATGCTTGGAAAACTGCTGTGCACAGGC	852
Qy	316	LeuValGlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAsp	335
Db	853	TTGGTCCAAAACATCCCGAACACCATTTAAGCAAGCTGATCAAGAAAGATTTTCAGGAT	912
Qy	336	MetLeuAsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCys	355
Db	913	ATGCTGAACATCTCTCAGCATCAATCGGTGAAAGCAATGTCGCCGAGAAATTCAGGTGC	972
Qy	356	PheArgHisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGly	375

```
QY 141 AspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSerTyrLysTyr 160
Db 256 GACAGCTTCAAGCTCTTGAATTCAGAACAGAACACACCTCTCTAAAGTCTTACAAATAC 315
QY 161 IleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSerTyrTyrGlu 180
Db 316 ATTAAGAGTCTGTTGAGAACGACATCAAGTTGCGCCAGGAAGAAATAGCTACTATGAG 375
QY 181 LysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValIleLysGluGlu 200
Db 376 AAAGTCTCGCTAAATACAAAGGACGACTTGGAAAGCATTAAGAAAGTAATCAAGAGAG 435
QY 201 LysGluLysPheProSerProProThrThrProProSerProAlaLysThrAspGlu 220
Db 436 AAGGAAAAGTTTCGAGCTCTCACCCACACACTCCCTCCCTGCAAGACCCGACGAG 495
QY 221 GlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrAsnAsn 240
Db 496 CAGAAAAAGAAAGTAAGTTCTTCATTCCTCACCAACATCGAAACTCTATATAACAAC 555
QY 241 LeuValAsnLysIleAspAspTyrTyrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsn 260
Db 556 CTGGTGAACAAGATTCATGACTACTTAACTTCAAGCGGAAATTAATGACTGTAAAC 615
QY 261 ValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLysAlaIleAsp 280
Db 616 GTCGAAAAGGATCAAGCCACGTTAAGATCACCAAGCTTTCCGATCTCAAAAGCCATCGAC 675
QY 281 AspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLysLeuIleAsn 300
Db 676 GATAAGATTGACTGTTTAAGAACCAACAACGATTTCCGACGCAATCAAAAGTTGATCAAC 735
QY 301 AspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuValGlnAsnPhe 320
Db 736 GAGGACTACTAAGAAACACATGCTTGAAAACTGCTGTCGACAGGCTTGGTCCAAAACCTC 795
QY 321 ProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeuAsnIleSer 340
Db 796 CCGAACACCATATAAGCAAGCTGTAGCGAAGAAAGTTTCAGGATATGCTGAACATCTCT 855
QY 341 GlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArgHisLeuAsp 360
Db 856 GAGCATCAATGCGTGAAGAACCAATGTCGCGAGAATTCAGGTTGCTTCGCGCACTTAGAC 915
QY 361 GluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLysCysValGlu 380
Db 916 GAAAGGAGGAGTGAATTAATGCTGCTGAATTTATAACAGAGAGAGCAAGTCGCTAGAG 975
QY 381 AsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAlaThrCysThr 400
Db 976 AATCCCTAACCCACCTGTACGAAAAATAACGGTGGCTGCGATGCTGACGCTAAGTGATCC 1035
QY 401 GluLysAspSerClySerArgLysLysIleThrCysGluCysThrLysProAspSer 420
Db 1036 GAGAGGACAGCGGTTCCCAATGCCAGAAAAATAACTTGGCAATGCACGAGCCCGATAGT 1095
QY 421 TyrProLeuPheAspGlyIlePheCysSer 430
Db 1096 TACCCTCTCTTCGACGGTATCTTCTGCTCC 1125
```

RESULT 3

```
US-10-098-514-1
; Sequence 1, Application US/10098514
; Publication No. US20020194648A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Sandra P
; APPLICANT: Christopher, David A
; APPLICANT: Vine, Benjamin
; APPLICANT: Su, Wei-wen
; APPLICANT: Bugos, Robert
; TITLE OF INVENTION: PLASMIDIUM FALCIPARUM MEROZOITE SURFACE PROTEIN-1 MALARIA PRODUCE
; TITLE OF INVENTION: TRANSGENIC PLANTS
; FILE REFERENCE: A-71339/RET/TAL/NBC
```

```
; CURRENT APPLICATION NUMBER: US/10/098,514
; CURRENT FILING DATE: 2002-08-06
; PRIOR APPLICATION NUMBER: US 09/500,376
; PRIOR FILING DATE: 2000-02-08
; PRIOR APPLICATION NUMBER: US 60/274,599
; PRIOR FILING DATE: 2001-03-09
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1149
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: synthetic
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: (1)..(24)
; OTHER INFORMATION:
US-10-098-514-1
Alignment Scores:
Pred. No.: 6,74e-167 Length: 1149
Score: 1900.00 Matches: 361
Percent Similarity: 98.38% Conservative: 4
Best Local Similarity: 97.30% Mismatches: 6
Query Match: 83.52% Indels: 0
DB: 9 Gaps: 0
```

US-10-057-531A-2 (1-431) x US-10-098-514-1 (1-1149)

```
QY 60 SerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIleTyrLeu 79
Db 13 AACATCTCTCAGTCACAACTCCTCAGTGCGTTCGAGAACGAGTACGACGTAATCTACCTA 72
QY 80 LysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIlePheThr 99
Db 73 AACCCCTTCGCGGTGTCTACCGTTTCATTAAGAAACAGATAGAAAAGATATTTTCACG 132
QY 100 PheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyrPheLeu 119
Db 133 TTCAACTCAACCTAAATGACATCTCACTCGGCTCAAGAAAGCGAAAATACTTCTCTC 192
QY 120 AspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyrIleIle 139
Db 193 GACGTGTGTAATCCGACCTTATGCAATTCACGACACATTAGCTCTTAACGAGTACATCATA 252
QY 140 GluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSerTyrLys 159
Db 253 GAGGACAGCTTCAAGCTCTTGAATTCAGAACAGAACACCCCTCTCTTAAAGTCTCTACAAA 312
QY 160 TyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSerTyrTyr 179
Db 313 TACATTAAGGAGTCTGTTGAGAACGACATCAAGTTCCGCCAGGAGGAATTAGCTACTAT 372
QY 180 GluLysValLeuAlaLysTyrLysAspLeuGluSerIleLysLysValIleLysGlu 199
Db 373 GAGAAAGTCTCGCTAAATACAAGGACGACTTGGAAGACATTAAGAAAGGTAATCAAGAA 432
QY 200 GluLysGluLysPheProSerSerProProThrThrProProSerProAlaLysThrAsp 219
Db 433 GAGAAAGGAAAGTTTCGAGCTCTCCACCCCACTCCCTCCCTGCAAGAACGAC 492
QY 220 GluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrAsn 239
Db 493 GAGCAGAAAAAGAAAGTAAGTTCTTCCATTCCTCACCAACATCGAACTCTATATAAC 552
QY 240 AsnLeuValAsnLysIleAspAspTyrTyrLeuIleAsnLeuLysAlaLysIleAsnAspCys 259
Db 553 AACCTGCTGAACAAGATTGATGACTACTTAATCAACTTGAAGCGGAAAAATTAATGACTGT 612
QY 260 AsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLysAlaIle 279
Db 613 AACGTCGAAAAGGATGAAGCCACGTTAAGATCACCAAGCTTTTCGATCTCTCAAGCCATC 672
```



RESULT 5  
US-10-087-464-58  
; Sequence 58, Application US/10087464  
; Publication No. US20030059436A1  
; GENERAL INFORMATION:  
; APPLICANT: Chishtil, Athar  
; APPLICANT: Oh, Steven  
; APPLICANT: Liu, David  
; APPLICANT: Goel, Vikas  
; APPLICANT: Li, Xuerong  
; TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses Thereof  
; CURRENT APPLICATION NUMBER: US/10/087,464  
; PRIOR FILING DATE: 2002-03-01  
; PRIOR APPLICATION NUMBER: US 06/272,930  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 58  
; LENGTH: 1131  
; TYPE: DNA  
; ORGANISM: Plasmodium falciparum  
US-10-087-464-58

Alignment Scores:  
Pred. No.: 6,95e-99 Length: 1131  
Score: 1165.50 Matches: 219  
Percent Similarity: 73.78% Conservative: 54  
Best Local Similarity: 59.19% Mismatches: 76  
Query Match: 51.23% Indels: 21  
DB: 9 Gaps: 2

US-10-057-531A-2 (1-431) x US-10-087-464-58 (1-1131)

QY	63	MetAspAnlleLeuSerGlyPheGluAsnGluTyrAspValIleTyrLeuLysProLeu	82
Db	19	ATTGATAACATCTTTCTAAATTCAGAAATGAATATGAGGTATTAATATAATAAACCTTTA	78
QY	83	AlaGlyValTyrArgSerLeuLysGlnIleGluLysAsnIlePheThrPheAsnLeu	102
Db	79	GCAGGTGTTTATAGAAGTTTAAAAACAATAGAAAATAACGGTTATGACATTTAATGTT	138
QY	103	AsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyrPheLeuAspValLeu	122
Db	139	AATGTTAAGGATTTTAAATCCACCATTATAAACGTGAAATTTCAAAATGTTTTA	198
QY	123	GluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyrIleIleGluAspSer	142
Db	199	CAATCAGATTTAAATCCCATATAAAGATTTAAACATCAAGTAATATGTTGCAAGATCCA	258
QY	143	PheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSerTyrLysTyrIleLys	162
Db	259	TATAAATTTCTTAATAAGAAAAAGAGATAAAATTCCTTAAGCAGTTATAATATATTAAG	318
QY	163	GluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSerTyrTyrGluLysVal	182
Db	319	GATTCATAGATACGGATATAAATTTGCCAATGATGTTCTTGCGATATATAAATATTA	378
QY	183	LeuAlaLysTyrLysAspAspLeuGluSerIleLysValIleLysGluLysGlu	202
Db	379	TCCGAAAAATAATAATCAGATTTAGATCAATTAATAAATAATATC-----	423
QY	203	LysPheProSerSerProProThrThrProProSerProAlaLysThrAspGluGlnLys	222
Db	424	-----AACGACAAACAAAGGT	438
QY	223	LysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrAsnAsnLeuVal	242
Db	439	GAATATGAGAATACCTTCCCCTTTTAAACAATATGAGACCTTATATAAACAGTTAAT	498
QY	243	AsnLysLeuAspTyrLeuIleAsnLysAlaLysIleAsnAspCysAsnValGlu	262
Db		::::::::::::::::::	

QY 263 LysAspGluAlaHisVallLysIleThrLysLeuSerAspLeuLysAlaIleAspAspLys 283  
||| ||||||| :|||: ||||| ||| |||||  
Db 559 AATCAACAGCTAGAGTTAAATAAAGAACTTAATTAATTAATAAAACAATTCAGACAAA 618  
: ||||| :|||: ||||| :|||: |||||  
QY 283 IleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLysLeuIleAsnAspAsp 302  
::: ||||| :|||: ||||| :|||: |||||  
Db 619 TTGGCAGATTTTAAAAAAAATAACAATTCGTTGGAATTTGCTGATTTTATCAACAGATTAT 678  
: ||||| :|||: ||||| :|||: |||||  
QY 303 ThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal---GlnAsnPhePro 321  
: ||||| :|||: ||||| :|||: ||||| :|||: |||||  
Db 679 AACCATAAACTATTGACAAAGTTCCCTTAGTACAGTATGGTTTGTGAAATCTTGCT 738  
: ||||| :|||: ||||| :|||: ||||| :|||: |||||  
QY 322 AsnThrIleLeuSerLysLeuIleGluGlyLysPheGlnAspMetLeuAsnIleSerGln 341  
|||: ||||| :|||: ||||| :|||: ||||| :|||: |||||  
Db 739 AAAACCGTTTATCTAATTTACTTTGATGAAACTTGCAGGATATGTTAAACATTTACAA 798  
: ||||| :|||: ||||| :|||: ||||| :|||: |||||  
QY 342 HisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArgHisLeuAspGlu 361  
|||: ||||| :|||: ||||| :|||: ||||| :|||: |||||  
Db 799 CACCAATGCGTAAAAAAACAATTCACAAAATTCGATGTTTTCAGACATTTATAGATGA 858  
: ||||| :|||: ||||| :|||: ||||| :|||: |||||  
QY 362 ArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLysCysValGluAsn 381  
|||: ||||| :|||: ||||| :|||: ||||| :|||: |||||  
Db 859 AGAAGAAGATGTAATGTTTATTAAATTACAAACAAGGTGATAAATGTGTTGAAAT 918  
: ||||| :|||: ||||| :|||: ||||| :|||: |||||  
QY 382 ProAsnProThrCysAsnGluAsnAsnGlyCysAspAlaAspAlaThrCysThrGlu 401  
|||: ||||| :|||: ||||| :|||: ||||| :|||: |||||  
Db 919 CCAAAATCCTACTTGTACGAAAAATAATGGTGGATGTGATGCAGATGCCAAATGTACGAA 978  
: ||||| :|||: ||||| :|||: ||||| :|||: |||||  
QY 402 GluAspSerGlySerArgLysLysIleThrCysGluCysThrLysProAspSerTyr 421  
|||: ||||| :|||: ||||| :|||: ||||| :|||: |||||  
Db 979 GAAGATTCAGTAGCAACGGAAGAAATCACATGTGAATGTTACTAAACCTGATTCCTTAT 1038  
: ||||| :|||: ||||| :|||: ||||| :|||: |||||  
QY 422 ProLeuPheAspGlyIlePheCysSerSer 431  
|||: ||||| :|||: ||||| :|||: ||||| :|||: |||||  
Db 1039 CCACCTTTTCGATGTTATTTTCTGCAGTCC 1068  
: ||||| :|||: ||||| :|||: ||||| :|||: |||||  
RESULT 6  
US-10-087-464-54  
; Sequence 54, Application US/10087464  
; Publication No. US20030059436A1  
; GENERAL INFORMATION:  
; APPLICANT: Chishtil, Athar  
; APPLICANT: Oh, Steven  
; APPLICANT: Liu, David  
; APPLICANT: Goel, Vikas  
; APPLICANT: Li, Xuerong  
; TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and  
; FILE REFERENCE: S1237/7019  
; CURRENT APPLICATION NUMBER: US/10/087,464  
; CURRENT FILING DATE: 2002-03-01  
; PRIOR APPLICATION NUMBER: US 06/272,930  
; PRIOR FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 54  
; LENGTH: 1137  
; TYPE: DNA  
; ORGANISM: Plasmodium falciparum  
US-10-087-464-54

Alignment Scores:  
Pred. No.: 7e-99 Length: 1137  
Score: 1165.50 Matches: 219  
Percent Similarity: 73.78% Conservative: 54  
Best Local Similarity: 59.19% Mismatches: 76  
Query Match: 51.23% Indels: 21  
DB: 9 Gaps: 2

US-10-057-531A-2 (1-431) x US-10-087-464-54 (1-1137)

QY 63 MetAspAnlleLeuSerGlyPheGluAsnGluTyrAspValIleTyrLeuLysProLeu 82

Db 25 ATGTGATACATACTTCTTAAATGAAATGAATGAGGTTTATATATATAAACCCTTTA 84  
Qy 83 AlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIlePheThrPheAsnLeu 102  
Db 85 CGAGCTGTTATAGAAGTTTAAATAAACAATAGAAATAGCTTATGACATTAATGTT 144  
Qy 103 AsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyrPheLeuAspValLeu 122  
Db 145 AATGTTAAGGATATTTAAATTCACGATTAAATAAACGTTGAAATTTCAAAATCTTTTA 204  
Qy 123 GluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyrIleIleGluAspSer 142  
Db 205 GAATCAGATTAAATCCATATANAAGATTAAATCAATCAAGTAATATGTTGTCAAGATCCA 264  
Qy 143 PheLysLeuLeuAsnSerGlnLysAsnThrLeuLysSerTyrLysTyrIleLys 162  
Db 265 TATAAATTTCTTAATAAGAAAAGAGATAAATTCCTTAAGCAGTTATATATATTAAG 324  
Qy 163 GluSerValGluAsnAspIleLysPheAlaGlnGluLysIleSerTyrTyrGluLysVal 182  
Db 325 GATTCAATAGATACGATATATAATTTGCAAAATGATGTTCTGGATATATAAATAATTA 384  
Qy 183 LeuAlaLysTyrLysAspLeuSerIleLysLysValIleLysGluLysGlu 202  
Db 385 TCCGAAAATATAATCAGATTAGATTCAATTAATAAATAATATATC----- 429  
Qy 203 LysPheProSerSerProThrProThrProSerProAlaLysThrAspGluGlnLys 222  
Db 430 -----AAGCAACAACAAGGT 444  
Qy 223 LysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrAsnLeuVal 242  
Db 445 GAAATGAGAATACCTTTCCCTTTTAAACAATATGTAGACCTTATATAAACAAGTTAAT 504  
Qy 243 AsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsnValGlu 262  
Db 505 GATAAATGATTATTTGTAATTCATTAGAACGAAAGTTCTAAATATATACATATAG 564  
Qy 263 LysAspGluAlaHisValLysIleThrLysLeuSerAspLysAlaIleAspLys 282  
Db 565 AAATCAAAACGTAGAAAGTTAAATAAACAACCTTAATTAATTAATAAACAATTCAGACAAA 624  
Qy 283 IleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLysLeuIleAsnAsp 302  
Db 625 TTGGCAGATTTTAAATAAATAAATAAATTCGTTGGAATTTGCTGATTTATCAACAGATTAT 684  
Qy 303 ThrLysLysAspMetLeuGlyLysLeuSerThrGlyLeuVal---GlnAsnPhePro 321  
Db 685 AACCATAATACTATTGACAAAGTTCTTAACTACAGGTATGGTTTGGAAATCTTGCT 744  
Qy 322 AsnThrIleLeuSerLysLeuIleGluLysPheGlnAspMetLeuAsnIleSerGln 341  
Db 745 AAAACCGTTTATCTAATTTCTGATGGAACCTTGCAAGGTATGTTAAACATTCACAA 804  
Qy 342 HisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArgHisLeuAspGlu 361  
Db 805 CACCAATCGGTAAATAAACAATGTCACAAATTTCTGGATGTTTCAGACATTTAGATGAA 864  
Qy 362 ArgGluGluCysLysCysLeuAsnTyrLysGlnGluLysAspLysCysValGluAsn 381  
Db 865 AGACAGAAATGTAATGTTTAAATTAATAAACAAGAGGTGATGTTGTTGAAAT 924  
Qy 382 ProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAlaThrCysThrGlu 401  
Db 925 CCAATCCTACTTGAACGAAATAATGGTGGATGTTGATGCAGATGCCAATGTACCGAA 984  
Qy 402 GluAspSerGlySerSerArgLysLysIleThrCysGluCysThrLysProAspSerTyr 421  
Db 985 GAAGATTCAGTACACGGAAGAAATACATGTGAATGTACTAAACCTGATCTTAT 1044  
Qy 422 ProLeuPheAspGlyIlePheCysSerSer 431  
|||||

Db 1045 CCACCTTTTCGATGCTATTTTCTGCTGAGTTCC 1074  
RESULT 7  
US-10-082-018-1  
; Sequence 1, Application US/10082018  
; Patent No. US20020144299A1  
; GENERAL INFORMATION:  
; APPLICANT: CHEN, LI How  
; APPLICANT: MEADE, Henry  
; TITLE OF INVENTION: NOVEL MODIFIED MSP-1 NUCLEIC ACID SEQUENCES AND  
; TITLE OF INVENTION: METHODS FOR INCREASING mRNA LEVELS AND PROTEIN  
; TITLE OF INVENTION: EXPRESSIONS IN CELL SYSTEMS  
; FILE REFERENCE: 107.637.121A  
; CURRENT APPLICATION NUMBER: US/10/082.018  
; PRIOR FILING DATE: 2002-02-20  
; PRIOR APPLICATION NUMBER: US/09/175,684  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 1065  
; TYPE: DNA  
; ORGANISM: preferably, a bacterium, virus, or parasite  
US-10-082-018-1  
Alignment Scores:  
Pred. No.: 1.5e-98 Length: 1065  
Score: 1161.50 Matches: 218  
Percent Similarity: 73.71% Conservative: 54  
Best Local Similarity: 59.08% Mismatches: 76  
Query Match: 51.05% Indels: 21  
DB: 12 Gaps: 2  
US-10-057-531a-2 (1-431) x US-10-082-018-1 (1-1065)  
Qy 63 MetaspasIleLeuSerGlyPheGluAsnGluTyrAspValIleTyrLeuLysProLeu 82  
Db 19 ATCGATACATCTCTGCTCCAGATCGAGAACGAGTACGAGGTGCTGTACCTGAAGCGCTG 78  
Qy 83 AlaGlyValTyrArgSerLeuLysGlnIleGluLysAsnIlePheThrPheAsnLeu 102  
Db 79 GCAGGGTCTACCGAGCGCTGAAGAAGCAGCTGGAGAACACAGTGTACCTTCAACGTG 138  
Qy 103 AsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyrPheLeuAspValLeu 122  
Db 139 AACGTGAAGGATATCTCTGAACAGCGGTTCACAAAGCGGAGAACTTCAAGAAGCTGCTG 198  
Qy 123 GluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyrIleIleGluAspSer 142  
Db 199 GAGAGCGATCTGATCCCTTACAAAGGATCTGACACAGCACTACGTGGTCAAGGATCCC 258  
Qy 143 PheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLysSerTyrLysTyrIleLys 162  
Db 259 TACAAGTCTCTGACAGGAGAGAGAGATAGTTCCTGACAGCTTACACTACATCAAG 318  
Qy 163 GluSerValGluAsnAspIleLysPheAlaGlnGluLysIleSerTyrTyrGluLysVal 182  
Db 319 GATAGCATTTGATACCGATATCAACTTCGCCAACGATGCTCTGGGACTACTACAAGATCTCTG 378  
Qy 183 LeuAlaLysTyrLysAspLeuGluSerIleLysValIleLysGluLysGlu 202  
Db 379 TCCGAGAGATGACAGCGGATCTGGATTCAATCAAGAAGTACATC----- 423  
Qy 203 LysPheProSerSerProThrProThrProSerProAlaLysThrAspGluGlnLys 222  
Db 424 -----AAGCAACAACGGA 438  
Qy 223 LysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrAsnLeuVal 242  
Db 439 GAGAACGAGAGATGATCTGCTTCTGCAACATCATCGAGACCTGTACAGACCGTCAAC 498  
Qy 243 AsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsnValGlu 262  
|||||







```

Qy 398 ThrCysThrGluGluAspSerGlySerArgLysLysLleThrCysGluCysThrLys 417
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 181 AAATGTACCGAAGAAGATTTCAGGTAGCAACGGAAGAAAAATCACATGTGAATGTACTAAA 240

Qy 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 241 CCTGATTCTTATCCACTTTTCGGATGGTATTTTCGCAGTCC 282

RESULT 12
US-09-134-333-4
: Sequence 4, Application US/09134333
: Patent No. US20020076403A1
: GENERAL INFORMATION:
: APPLICANT: LONGACRE-ANDRE, SHIRLEY
: APPLICANT: ROTH, CHARLES
: APPLICANT: NATO, FARIDABANO
: APPLICANT: BARNWELL, JOHN
: APPLICANT: MENDIS, KAMINI
: TITLE OF INVENTION: RECOMBINANT PROTEIN CONTAINING A C-TERMINAL FRAGMENT
: TITLE OF INVENTION: PLASMODIUM MSP-1
: FILE REFERENCE: 0660-0135-0XCIP
: CURRENT APPLICATION NUMBER: US/09/134.333
: CURRENT FILING DATE: 1999-04-18
: EARLIER APPLICATION NUMBER: PCT/FR97/00290
: EARLIER FILING DATE: 1997-02-14
: EARLIER APPLICATION NUMBER: FR96/01822
: EARLIER FILING DATE: 1996-02-14
: NUMBER OF SEQ ID NOS: 14
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 4
: LENGTH: 354
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence:SYNTHETIC
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1)..(354)
US-09-134-333-4

Alignment Scores:
Pred. No.: 1.87e-40 Length: 354
Score: 527.00 Matches: 91
Percent Similarity: 97.87% Conservative: 1
Best Local Similarity: 96.81% Mismatches: 2
Query Match: 23.16% Indels: 0
DB: 10 Gaps: 0

US-10-057-531A-2 (1-431) x US-09-134-333-4 (1-354)
Qy 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 7 AACATCTGCGAGCACCAATGCTGAAAAACAATGTCCCGAGAACTCTGGCTGTTTCAGA 66

Qy 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 67 CACTTGGACGAGAGAGAGAGTGTAAATGTCTGCTGAACATACAACAGAGGGCGACAA 124

Qy 378 CysValGluAsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAla 397
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 127 TGC GTGGAGAACCCCAACCCCACTGTAAACGAGAACAAACGGCGGTGTGACGACAGACGC 187

Qy 398 ThrCysThrGluGluAspSerGlySerArgLysLysLleThrCysGluCysThrLys 417
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 187 AAATGCACCGAGGAGGACTCGGGCAGCAACGGCAGAAAAATACAGTGTGAGTGACCAA 240

Qy 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 247 CCCGACTCGTACCGCGTGTTCAGCGGCATCTTCGCAGCTCC 288

RESULT 13
US-09-134-333-3
: Sequence 3, Application US/09134333

```

```
; Patent No. US20020076403A1
; GENERAL INFORMATION:
; APPLICANT: LONGACRE-ANDRE, SHIRLEY
; APPLICANT: ROTH, CHARLES
; APPLICANT: NATO, FARIDABANO
; APPLICANT: BARNWELL, JOHN
; APPLICANT: MENDIS, KAMINI
; TITLE OF INVENTION: RECOMBINANT PROTEIN CONTAINING A C-TERMINAL FRAGMENT OF
; FILE REFERENCE: 0660-0135-0XCIP
; CURRENT APPLICATION NUMBER: US/09/134,333
; EARLIER FILING DATE: 1999-04-18
; EARLIER APPLICATION NUMBER: PCT/FR97/00290
; EARLIER FILING DATE: 1997-02-14
; EARLIER APPLICATION NUMBER: FR96/01822
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 279
; TYPE: DNA
; ORGANISM: Plasmodium falciparum
US-09-134-333-3

Alignment Scores:
Pred. No.: 3,166-40 Length: 279
Score: 523.00 Matches: 90
Percent Similarity: 97.85% Conservative: 1
Best Local Similarity: 96.77% Mismatches: 2
Query Match: 22.99% Indels: 0
DB: 10 Gaps: 0

US-10-057-531a-2 (1-431) x US-09-134-333-3 (1-279)

Qy 338 AsnIleSerGlnHisGlnCysValIysLysGlnCysProGluAsnSerGlyCysPheArg 357
Db 1 AACATTCACAAACCAATCGTGAATAAACAATGTCACGAAATTCGGATGTTTCA 60
Qy 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377
Db 61 CATTTAGATGAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
Qy 378 CysValGluAsnProAsnProThrCysAsnGluAsnGlyGlyCysAspAlaAspAla 397
Db 121 TGTGTTGAATCCAAATCCTACTGTAAACGAAATAATATGTTGATGATGATGCC 180
Qy 398 ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417
Db 181 AAATGTACCGAAGAGATTCAGGTAGCAACGGAAGAAATAATCAGATGTGAATGTA 240
Qy 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSer 430
Db 241 CTGATCTTATCCACTTTTCGATGGTATTTCTGCAGT 279

RESULT 14
US-09-134-333-1
; Sequence 1, Application US/09134333
; Patent No. US20020076403A1
; GENERAL INFORMATION:
; APPLICANT: LONGACRE-ANDRE, SHIRLEY
; APPLICANT: ROTH, CHARLES
; APPLICANT: NATO, FARIDABANO
; APPLICANT: BARNWELL, JOHN
; APPLICANT: MENDIS, KAMINI
; TITLE OF INVENTION: RECOMBINANT PROTEIN CONTAINING A C-TERMINAL FRAGMENT OF
; FILE REFERENCE: 0660-0135-0XCIP
; CURRENT APPLICATION NUMBER: US/09/134,333
; EARLIER FILING DATE: 1999-04-18
; EARLIER APPLICATION NUMBER: PCT/FR97/00290
; EARLIER FILING DATE: 1997-02-14
; EARLIER APPLICATION NUMBER: FR96/01822
; EARLIER FILING DATE: 1996-02-14
```

```
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 291
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(291)
US-09-134-333-1

Alignment Scores:
Pred. No.: 3,356-40 Length: 291
Score: 523.00 Matches: 90
Percent Similarity: 97.85% Conservative: 1
Best Local Similarity: 96.77% Mismatches: 2
Query Match: 22.99% Indels: 0
DB: 10 Gaps: 0

US-10-057-531a-2 (1-431) x US-09-134-333-1 (1-291)

Qy 338 AsnIleSerGlnHisGlnCysValIysLysGlnCysProGluAsnSerGlyCysPheArg 357
Db 7 AACATCTCGACACCAATCGGTGAATAAACAATGTCACGAACTCTGGCTGTTTCA 66
Qy 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377
Db 67 CACTTGGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 126
Qy 378 CysValGluAsnProAsnProThrCysAsnGluAsnGlyGlyCysAspAlaAspAla 397
Db 127 TCGCTGGAGAACCCCAACCCGACCTGTAAACGAGAAACACGCGGCTGTGACGAC 186
Qy 398 ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417
Db 187 AAATGCACCGAGGAGGAGTCTGGGACGACCAACGCGCAAGAAATACACGTGTG 246
Qy 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSer 430
Db 247 CCCGACTCGTACCCGCTGTTCTGACGCGCATCTTCTGCAGC 285

RESULT 15
US-09-134-333-7
; Sequence 7, Application US/09134333
; Patent No. US20020076403A1
; GENERAL INFORMATION:
; APPLICANT: LONGACRE-ANDRE, SHIRLEY
; APPLICANT: ROTH, CHARLES
; APPLICANT: NATO, FARIDABANO
; APPLICANT: BARNWELL, JOHN
; APPLICANT: MENDIS, KAMINI
; TITLE OF INVENTION: RECOMBINANT PROTEIN CONTAINING A C-TERMINAL FRAGMENT OF
; FILE REFERENCE: 0660-0135-0XCIP
; CURRENT APPLICATION NUMBER: US/09/134,333
; EARLIER FILING DATE: 1999-04-18
; EARLIER APPLICATION NUMBER: PCT/FR97/00290
; EARLIER FILING DATE: 1997-02-14
; EARLIER APPLICATION NUMBER: FR96/01822
; EARLIER FILING DATE: 1996-02-14
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 387
; TYPE: DNA
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(387)
US-09-134-333-7
```





```
QY 184 AlaLysTyrLysAspLeuGluSerIleLysLysValIleLysGluLys----- 201
Db 1147 CCGAGCTAT-----GGCTAGAGTCCATTTCAGAGGCTAATTGCCAGGTCCTATTCT 1200
QY 202 ---GluLysPheProSer-----SerProThr-Th 211
Db 1201 CTAAATAATGCCAATCAAGAGAAACATTGTCAATCTCTGGAGGCCAGCTTGACTTAC 1260
QY 211 rProProSerProAla-----LysThrAspGluGlnLysLysGluSerLysPheLeuPr 229
Db 1261 CCCAGCCACTGCTGCTTTTGAACACTTGCACAAACAAAGACAGAAATTTTCACATCC 1320
QY 229 oPheLeuThrAsnIleGluThrLeuTyrAsnAsnLeu 241
Db 1321 ATTCTGT-----AAACACTTTTCCAAACAATGTG 1348

RESULT 19
US-10-046-583A-6
; Sequence 6, Application US/10046583A
; Patent No. US20020168743A1
; GENERAL INFORMATION:
; APPLICANT: Rice, John
; APPLICANT: Kloti, Andreas
; APPLICANT: Crawford, John
; APPLICANT: Lanning, Beth
; APPLICANT: Stewart, Sandy
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
; TITLE OF INVENTION: IDENTIFICATION OF MODULATORS OF DEOXYXYLULOSE 5-PHOSPHATE
; FILE REFERENCE: 2037 US Divisional
; CURRENT APPLICATION NUMBER: US/10/046.583A
; PRIOR FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: 09/626,589
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 2472
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Nucleotides 1-495 encode the thioredoxin sequence
; OTHER INFORMATION: found in the vector PET32 supplied by No. US20020168743A1agen.
; OTHER INFORMATION: Nucleotides 496-2472 represent the tDXPS cDNA
; OTHER INFORMATION: sequence from Arabidopsis thaliana.
US-10-046-583A-6

Alignment Scores:
Pred. No.: 2,88e-16 Length: 2472
Score: 278.50 Matches: 76
Percent Similarity: 49.75% Conservative: 25
Best Local Similarity: 37.44% Mismatches: 49
Query Match: 12.24% Indels: 53
DB: 9 Gaps: 7

US-10-057-531A-2 (1-431) x US-10-046-583A-6 (1-2472)
QY 1 MethHisHisHisHisSerSerGlyLeuValProArgGlySerGlyMetLysGlu 20
Db 346 ATGCACCATCATCATCATCTTCTGCTGCTGGTCCACGCGTCTCTGGTATGAAGAA 405
QY 21 ThrAlaAlaLysPheGluArgGlnHisMetAspSerProAspLeuGlyThrAsp 40
Db 406 ACCGCTGCTGCTAAATTCGAACGCCACACATGGACAGCCAGATCTGGTACCGACGAC 465
QY 41 AspAspLysAlaMetAlaAspIleGlySerIleGluGlyThrMetAlaIleSer 60
Db 466 GAGCAAGCGCCATGCTGCTATCGGATCCGCTTCA----- 501
QY 61 ValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIleTyrLeuLys 80
Db 501 ----- 501
```

```
QY 81 ProLeuAla-----GlyValTyrArgSerLeuLysGlnIleGluLysAsnIlePhe 98
Db 502 ---CTTGCAGAGAGGGTGAATATTATTCAC---AACAGACCACCACTCCATTACTTGAC 555
QY 99 ThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyrPhe 118
Db 556 ACTATTAACTACCAATCCAC---ATGAAAAATCTTTCTGTCAAGGAAGTGAACAACATT 612
QY 119 LeuAspValLeuLysSerAspLeuMetGlnPheLys-----HisIle 132
Db 613 TCTGATGACGTGAGATCAGACGTGATC---TTTAATGTCTCGAAAAACCGGTGGACATTG 669
QY 133 SerSerAsnGluTyrIleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsn 152
Db 670 GGGTCAAGTCTTGGTGTCTGGGAG----- 693
QY 153 ThrLeuLeuLysSerTyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAla 172
Db 694 ---CTTACTGTGGCTCTTCAATACATTTTCAATACTCCACAGACAAGATTCTTTGGGAT 750
QY 173 GlnGluGlyIleSerTyrGluLysValLeuAlaLysTyrLysAspLeuGluSer 192
Db 751 GTTGGTCATCAGTCTTATCTCTCATAGATTCTTACTGGGAGAGAGGAAGATGCCTACA 810
QY 193 IleLysLys 195
Db 811 ATGAGGCAA 819

RESULT 20
US-10-012-896-1010
; Sequence 1010, Application US/10012896
; Publication No. US20020183251A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Devin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; APPLICANT: Wantanabe, Yoshihiro
; APPLICANT: Meagher, Madeleine Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C27
; CURRENT APPLICATION NUMBER: US/10/012,896
; CURRENT FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 1011
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1010
; LENGTH: 1065
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-012-896-1010

Alignment Scores:
Pred. No.: 2,71e-15 Length: 1065
Score: 262.50 Matches: 60
```



Qy	270	IleThrLysLeuSerAspLeuLysAlaIleAspAspLysIleAspLeuPheLysAsnPro	289
Db	1231	---AATCTAATTAAGGTTTACGAAAGGTCGAAGATTCAATAGGT-----AATAATGAA	1281
Qy	290	TyrAspPheGluAlaIleLysLysLeuIleAsn-----	300
Db	1282	AGTGATTACGAGAAATTTATCGCAATAAATATCGATACATAACATAAATAATCAAAATT	1341
Qy	301	-----AspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGly	315
Db	1342	AATGTAATTAAGAAATGAAAGAGCTAAAGATGAATTAGATATAACTATTATTA-----	1392
Qy	316	LeuValGlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAsp	335
Db	1393	-----GGTAGTAACCAAGAG	1407
Qy	336	MetLeuAsn---IleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGly	354
Db	1408	TTAGAGAATCAAAATTAATGAAGAAAAAACAATATATGAAGAATTCGAAATAAAAA-----	1461
Qy	355	CysPheArgHisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGlu	374
Db	1462	-----TTAGATCGCTATGATAAAATCAAAATTTGGACTTAAATGATATAAAGAAGAAGC	1509
Qy	375	-----GlyAspLysCysValGluAsnProAsn	383
Db	1510	TTTATAAGTCAAAATTAATCTGCTGTAAGATTGGAGATCAATGT-----	1554
Qy	384	ProThrCys---AsnGlu	388
Db	1555	CCGATATGTGTAATGAA	1572

RESULT 23

US-09-815-242-8043

; Sequence 8043, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Karl L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; TITLE OF INVENTION: Prokaryotes

; FILE REFERENCE: ELITRA.011a

; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 8043

; LENGTH: 3030

; TYPE: DNA

; ORGANISM: Staphylococcus aureus

; FEATURE:

; NAME/KEY: CDS





Db 1108 TTCTTTAAACAAGGTATCCCATCGAAGAAATTCCTCAATGCACCTT-----ATC 1158  
QY 127 MetGlnPheLysHisIleSerSerAsnGluTyrIleIleGluAspSerPheLysLeuLeu 146  
Db 1159 AGACGCTTGCACACTGCTGTGTTAAATA-----GATATCTGT 1197  
QY 147 AsnSerGluGlnLysAsnThrLeuLeuLysSerTyrLysTyrIleLysGluSerValGlu 166  
Db 1198 AATTCAGAGACCAAAAGGATTACTCAAG-----AAGGTTAAA 1236  
QY 167 AsnAspIleLysPheAlaGlnGluGlyIleSerTyrTyr----- 179  
Db 1237 AATGCCATTTTCACCTACAGCAGAGAGATCTCTGGTTATATGCGAATGATGCTCATGAA 1296  
QY 180 -----GluLysValLeuAlaLysTyrLysAspAspLeuSerIleLysValIle 197  
Db 1297 TTTTAAAGTCAGTGTGTGGACAGCTGAAGAAGATATGAAAAATTAATRAAACTGG 1356  
QY 198 LysGluGluLysGluLysPheProSerSerProProThrThrProProSerProAlaLys 217  
Db 1357 AAGACTGAACCTGTTTCTGTGAGAGAAAATTCACAGATATTTTCAGCTACACAGCATAC 1416  
QY 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGlu---Thr 236  
Db 1417 ACT-----TGCCCTGTTTATTACTAATTTGGAGTTTGAG 1449  
QY 237 LeuTyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIle 256  
Db 1450 GTTCAGCACTCCATCATTTGAAGCATGTGGAGATTTATCCCAAAAGACAGACGTTT 1509  
QY 257 AsnAspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeu 276  
Db 1510 AATGACCTCTCTATTGACCTTCTCGTAGGAAAAAACCACTCCCTCCT----- 1557  
QY 277 LysAlaIleAspLysIleAspLeuPheLysAsnProTyrAspPheGlu---AlaIle 295  
Db 1558 CGTTCAATTCAGATTCTCTTGATCTTTCTTTAGGCCCGAAGAACTGGAGTATCTTGT 1617  
QY 296 LysLys-----LeuIleAsnAspAspThrLysLysLysAspMetLeuGly 309  
Db 1618 GAGAGTGTGGTGGAGTGTCTCTGTTCAGGCAC-----AAATTTAACAGCTTCTCT 1671  
QY 310 LysLeuLeuSerThrGlyLeuValGlnAsnPheProAsnThrIleIleSerLysLeuIle 329  
Db 1672 AGGTCCTCATCTTCCATTTGAACAGATATAGCTTCAATGTGGCTCTCTCG-----CTT 1725  
QY 330 GluGlyLysPheGlnAspMetLeuAsnIleSerGlnHisGlnCysValLysLysGlnCys 349  
Db 1726 AACAATAAGATTGGGCAGCAAGTCATCTTCAAGATACCTGACCCTGTCTCATCTCAATTGC 1785  
QY 350 ProGluAsnSer-----GlyCysPheArgHisLeuAspGluArgGlu 363  
Db 1786 ACTGAAATAACAAACCAACCTTTTACCCTTGGTGGAGTGCACATATGCGCAATGCTAGA 1845  
QY 364 GluCysLysCysLeuLeuAsnTyrLysGlnGluLysCysValGluAsnProAsn 383  
Db 1846 CCATTGAAAGCC-----TCTCAAAATGCTGAATCTCTGCATCCACAGCCCTTCT 1893  
QY 384 ---ProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAlaThrCysThrGluGlu 402  
Db 1894 ACACCTTCAAGAAATTCACCTTCAATCCAAAGAGCTCTTGGCTTTATGCCCTTGATTCA 1953  
QY 403 AspSerGlySerArgLysLys 410  
Db 1954 GACAGTGAGTGAGCTAAACGT 1977

RESULT 25

US-09-974-300-2660  
; Sequence 2660, Application US/09974300  
; Patent No. US20020146721A1  
; GENERAL INFORMATION:  
; APPLICANT: Berka, Randy M.  
; APPLICANT: Clausen, Ib Groth

; TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
; FILE REFERENCE: Expression  
; CURRENT APPLICATION NUMBER: US/09/974,300  
; CURRENT FILING DATE: 2001-10-05  
; PRIOR APPLICATION NUMBER: 09/680,598  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: 60/279,526  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 8481  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2660  
; LENGTH: 3057  
; TYPE: DNA  
; ORGANISM: Bacillus licheniformis  
US-09-974-300-2660

Alignment Scores:  
Pred. No.: 0.00492 Length: 3057  
Score: 137.00 Matches: 109  
Percent Similarity: 36.97% Conservative: 57  
Best Local Similarity: 24.28% Mismatches: 163  
Query Match: 6.02% Indels: 120  
DB: 10 Gaps: 24

US-10-057-531A-2 (1-431) x US-09-974-300-2660 (1-3057)

QY 8 SerSerGlyLeuValProArgGlySerGlyMetLysGluThrAlaAlaLysPheGlu 27  
Db 1303 TCCTCAGGAATGAGCGCACATTCAGGAGATAAAGACCATCTGCGGAGAGACCCGAAT 1362  
QY 28 ArgGlnHisMetAspSerProAspLeuGlyThrAsp---AspAspAspLysAlaMetAla 46  
Db 1363 CAAGGACAAGAACCTTCCCTCGGAACCGGGAAACGGTCAGCAAGCGCGCGCAACT 1422  
QY 47 AspIleGlySerIleGluGlyArgGlyThrMetAlaIleSerValThrMetAspAsnIle 66  
Db 1423 CCTTCAGATGAGATAAATGATCAAGGGACG----- 1452  
QY 67 LeuSerGlyPheGluAsnGluTyrAspValIleTyrLeuLysProLeuAlaGlyValTyr 86  
Db 1453 -----CGTGAAGATGATCAA-----CAGCTGAACCT----- 1479  
QY 87 ArgSerLeuLysLysGlnIleGluLysAsnIlePheThrPheAsnLeuAsnLeuAsnAsp 106  
Db 1480 -----CGAAACGAACAACCGGAAGATGAG-----AATGCTAAGTTAAAGAA 1521  
QY 107 IleLeuAsnSerArgLeuLysLysArgLysTyrPheLeuAspValLeuGluSerAspLeu 126  
Db 1522 CAATTGAACGAAGCGCGCAGCAAGG-----CTTGAAAAATTTGGAAGAAGATCTT 1569  
QY 127 -----MetGlnPheLysHisIleSerSerAsnGluTyrIleIleGluAspSerPheLys 144  
Db 1570 CGTGCTAAACGAGACACTCATACGAAGAGCTGGAAAAACAGATTGAAATTAATAAG 1629  
QY 145 LeuLeuAsnSerGluGlnLysAsn-----ThrLeuLeuLysSerTyrLysTyrIleLys 162  
Db 1630 AAAATTAGTAGTCTGAAGAAGAAACATTAAAGCATTAAGCATGAATCAAAATTTGAAG 1689  
QY 163 GluSerValGluAsn-----AspIleLysPheAla 172  
Db 1690 GTGGATATTGAGATCTCAATAAAAAATTTTGTATGTAATCTATGATGATTTAAAAAATTT 1749  
QY 173 GlnGluGlyIleSerTyrTyrGluLysValIleAlaLysTyrLysAspAspLeuGluSer 192  
Db 1750 GAAATGGCATATTAGAGTAT-----CTCCGAAAGATAGTAGTCGGAAGAGAAGC 1800  
QY 193 IleLysLysValIleLysGluGluLysPheProSerSerProProThrThrPro 212  
Db 1801 TTAGCAAGGCTTTTGAG----- 1818  
QY 213 ProSerProAlaLysThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThr 232  
Db 213 ProSerProAlaLysThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThr 232



```
Db 1720 AAATCCTTGCAGGCTGACACC-----ACCAACACTGAC----- 1752
Qy 225 SerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrAsnAsnLeuValAsnLys 244
Db 1753 -----ACTGCCTTGACAACCTTTGGAGGAGGCCCTTGCAGAGAAA 1791
Qy 245 IleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsnValGluLysAsp 264
Db 1792 ---GAGCGGCAATTGAACGCTTAAGAGGAGGAGGAGAGAGATGAGCGAGAGAAACAA 1848
Qy 265 GluAlaHisValLysIleThrLysLeuSerAspLeuLysAlaIleAspAspLysIleAsp 284
Db 1849 GAGGAATTTGATACTACAAAAA-----GATCTTAAAGACTTTGAAGGAAAAAGTCAAGC 1902
Qy 285 LeuPheLysAsnProTyrAspPheGluAlaIleLysLysLeuIleAsnAspAspThrLys 304
Db 1903 CTGTTGCAAGCGGACCTTTCAGAGAAAGAGGCTTCACTTTG-----GATCTG 1950
Qy 305 LysAspMetLeuGlyLysLeuLeuSerThrGlyLeuValGlnAsnPheProAsnThrIle 324
Db 1951 AAAGAGCATGCTTCTCTGGCATCTCAGGACTGAAAGAGGACTCAGGCTTAAGACA 2010
Qy 325 IleSerLysLeuIleGluGlyLysPheGlnAspMetLeuAsnIleSerGln----- 341
Db 2011 CTAGAGATTGCTTTGGAGCAGAGAGGAGGAGTGTCTGAAATTTGAATACAAATTTGAAA 2070
Qy 342 -----HisGlnCysVal-----LysLysGlnCysProGluAsnSerGlyCysPheArg 357
Db 2071 AAGCCATGAGGAGGAGGAGTGAAGAGCCAGAGCCAGTCCAGAGATGATGACCGCAATACAG 2130
Qy 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGly----- 375
Db 2131 CACTTG---GAGAGAGAG-----ATCACCAGGTACAAAGATGATCTAGCAAG 2175
Qy 376 -----AspLysCysValGluAsnProAsnProThrCysAsnGluAsnAsn 390
Db 2176 GCCCAGGACAGAAGTTGATCGACTCTTAGAAATCTTGAAGGAGGTGGAATAATGAGAAGAA 2235
Qy 391 GlyGlyCysAspAlaAspAlaThrCysThrGluGluAspSerGlySerSerArgLys 409
Db 2236 -----GACAAGATAGAAGATAGCTGAGTTGGAAGATCTCACCTCAAGGCAA 2283

RESULT 27
US-10-267-311-20
: Sequence 20, Application US/10267311
: Publication No. US20030050469A1
: GENERAL INFORMATION:
: APPLICANT: Siegel, Marvin
: APPLICANT: Chu, N. Randall
: APPLICANT: Mizzen, Lee A.
: TITLE OF INVENTION: INDUCTION OF A THI-LIKE RESPONSE IN VITRO
: FILE REFERENCE: 12071/002001
: CURRENT APPLICATION NUMBER: US/10/267,311
: PRIOR APPLICATION DATE: 2002-10-09
: PRIOR FILING DATE: 2000-07-10
: PRIOR APPLICATION NUMBER: US 60/143,757
: PRIOR FILING DATE: 1999-07-08
: NUMBER OF SEQ ID NOS: 55
: SOFTWARE: FastSEQ for Windows Version 4.0
: SEQ ID NO 20
: LENGTH: 2847
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: fusion sequence
: NAME/KEY: CDS
: LOCATION: (1)...(2844)
US-10-267-311-20
```

Alignment Scores:

```
Pred. No.: 0.0144 Length: 2847
Score: 131.50 Matches: 82
Percent Similarity: 35.12% Conservative: 62
Best Local Similarity: 20.00% Mismatches: 123
Query Match: 5.78% Indels: 143
DB: 9 Gaps: 16

US-10-057-531A-2 (1-431) x US-10-267-311-20 (1-2847)
Qy 2 HisHisHisHisHisHisSerSerGlyLeuValProArgGlySerGlyMetLysGluThr 21
Db 13 CATCATCATCATCATCACAGCGCGCTGGTGGCCGCGGAGCCATATGCCAAGACA 72
Qy 22 AlaAlaAlaLysPheGluArgGlnHisMetAspSerProAspLeuGlyThrAspAsp 41
Db 73 ATTCCGTACGACGAGAGGCCCGTCGC-----GGCTCGAGCGGGGC 114
Qy 42 AspLysAlaMetAlaAspIleGlySerIle-----GluGlyArgGlyThrMet 57
Db 115 TTGAACGCCCTTCGCCGATCGCGTAAAGGTGACATTGGGCCCGCCAGGGCCGCAACGTCGTC 174
Qy 58 -----AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGlu 71
Db 175 CTGGAAGAAGAGTGGGTGGCCCGATCACCAACGATGTTGTCTCATCCCAAGAG 234
Qy 72 AsnGluTyrAspValIleTyrLeuLys----- 80
Db 235 ATCGAGCTGGAGGATCCGTACGAGAGATCGGCCCGGAGTGGTCAAGAGGTAGCCAAAG 294
Qy 80 ----- 80
Db 295 AAGACCGATGACGTGCGCGGTGACGGGACACACGACGCGCCACCGTCTGGCCGAGCGTTG 354
Qy 81 -----ProLeuAlaGlyValTyrArg 87
Db 355 GTTCGCGAGGCGCTCGCAACGTCGCGCGCGCGCCGCGCCGCTC---GGTCTCAACGC 411
Qy 88 SerLeuLysLysGlnIleGluLysAsnIlePheThrPhe----- 100
Db 412 GGCATCGAAGAGCGCGTGAGAGAGTCCACGAGACCTGCTCAAGGGCGCCCAAGAGGTC 471
Qy 101 -----AsnLeuAsnLeuAsn 105
Db 472 GAGACCAAGAGACAGATTGCGGCCACCGACGAGTATTCGGCGGGTGACAGTCCATCGGT 531
Qy 106 AspIleLeuAsnSerArgLeuLysLys-----ArgLysTyrPheLeuAspValLeuGlu 123
Db 532 GACCTGATCGCGGAGCGCATGACAAGTGGGCAACGAGGGGCTCATCCGCTCGAGGAG 591
Qy 124 SerAspLeu-----MetGlnPheLys-----His 131
Db 592 TCCAACACCTTTGGGCTGCGAGCTCGAGCTACCGAGGGTATGCGGTTCCGACAGGGCTAC 651
Qy 132 IleSerSer-----AsnGluTyrIleIleGluAspSerPhe 143
Db 652 ATCTCGGGGTACTTCGTGACCGACCGCGGAGCGTTCAGAGAGCGGCTCTCGAGGACCCCTAC 711
Qy 144 LysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSerTyrLysTyrIleLysGlu 163
Db 712 ATCTCTGTGTACCTCAAGGTGTCCACTGTC----- 744
Qy 164 SerValGluAsnAspIleLysPheAlaGlnGluIleSerTyrTyrGluLysValLeu 183
Db 745 -----AAGGATCTGCTGCCGCTGCTCGAAGAGGTATC 777
Qy 184 AlaLysTyrLys-----AspAspLeuGluSer-----IleLys 194
Db 778 GGAGCCGGTAAAGCGCGCTGCTGATCATCCCGAGGACGTCGAGGGCGGAGCGGTGTCCACC 837
Qy 195 LysValIleLysGluGluLysPheProSerSerProThrThrProProSer 214
Db 838 CTGCTGTCACAAAGATCCCGCGGACCTTCAAGTCG-----GTGGCGGTCAAGGCT 888
```



```
; LENGTH: 2341
; TYPE: DNA
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (966)...(2291)
US-09-881-752A-147

Alignment Scores:
Pred. No.: 0.0187 Length: 2341
Score: 129.00 Matches: 89
Percent Similarity: 41.58% Conservative: 64
Best Local Similarity: 24.18% Mismatches: 125
Query Match: 5.67% Indels: 91
DB: 10 Gaps: 19

US-10-057-531A-2 (1-431) x US-09-881-752A-147 (1-2341)
QY 85 ValTyrArgSerLeuLysLysGlnIleGluLysAsnIlePheThrPheAsnLeuAsnLeu 104
Db 562 GTAAAGAGAAGTTTAAAGAAAAATCCTTAAGGAATTTGAAAAACGATAAATAC-AATCAA 620
QY 105 AsnAspIleLeuAsn---SerArgLeuLysLysArg----- 115
Db 621 ACGCAAAATAGTAGGGTTAGAAAAATTTAAAGGAAAAAATTCAGATTCGTTTTGGTGAAAAAC 680
QY 116 LysTyrPheLeuAspValLeuGluSerAspLeuMetGlnPhe----- 129
Db 681 CAAACAGAATTGGCACTATTGGAATCCAAATTTAACAGATTTTCATTTTATTGAAATCAT 740
QY 130 -----LysHisIleSerSerAsnGluTyrIleIleGluAspSerPheLys 144
Db 741 TCTATTGGGAACAAAAATTTAGGAGAGTGGTGATCGACCCATTGCAGATTTAATAAAAA 800
QY 145 LeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSerTyrLysTyrIleLysGluSer 164
Db 801 AGATTAAAGCAAGAA-----GATTGGGTAGCTCAGGTAGAGATATATATAAAA----- 848
QY 165 ValGluAsnAspIle---LysPheAlaGlnGluIleSerTyrTyrGluLysValLeu 183
Db 849 ---GATAATAGTATATATGCGCTTTCTGTCAAAAA-----GAAACCATTTACC 890
QY 184 AlaLysTyrLysAspAspLeuGlu-----SerIleLysLysValIleLys 198
Db 891 GAAGAATTTAAAAACAACACTAGAACTTATTGATACAAAGTTATCAAGAATCTATTGAA 950
QY 199 GluGluLysGluLysPheProSerProThrThrProProSerProAlaLysThr 218
Db 951 ACGATCAAGGAAAGATGGAAGACTACGACGACGACGACGACGACGACGACGACGACGACG 1010
QY 219 Asp-----GluGlnLysLysGluSerLysPheLeuPheLeuThrAsn 233
Db 1011 GATAAGATTTGTAACACAGACAGAGAAATCAACAAACTAAATTTGGAC---ACAGAAAT 1067
QY 234 IleGluThrLeuTyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeu--- 252
Db 1068 TTGAAAAATAATTATGAAACTTTTGAGAAAGTAAATCAATCGGAATCAGCAAAAGATGCTT 1127
QY 253 ---LysAlaLysIleAsnAspCysAsnValGluLysAspGluAlaHisValLysIleThr 271
Db 1128 GATAAAGTAAAGAAATGACGAGAAATTTTAAGCTTGATAGCACTAAAACGAGATAGAC 1187
QY 272 LysLeuSerAspLeu---LysAlaIleAspAspLysIle-----AspLeuPhe 286
Db 1188 GCAATTTAAAGATTGTGATTTAAAAAGGCTAATGAGCAAAATGACCAATTAATGAGATGATA 1247
QY 287 LysAsnProTyrAspPheGluAlaIleLysLysLeuIleAsnAspThrLysLysAsp 306
Db 1248 AAG-----GATATTGAAAAACAGAAAGAGTTGTAAGGAACAACTTGGAAATTT 1298
QY 307 MetLeuGlyLysLeuLeuSerThrGlyLeuValGlnAsnPheProAsnThrIleIleSer 326
Db 1299 CTAGTCAATGAATTTAAAGT----- 1319
```

```
QY 327 LysLeuIleGluGlyLysPheGlnAspMetLeuAsnIleSerGlnHisGlnCysValLys 346
Db 1320 -----GATATACAGAAATATAAT-----AAA 1340
QY 347 LysGlnCysProGluAsnSerGlyCysPheArgHisLeuAspGlu----- 361
Db 1341 AAGTATTGCGGTTTGGAGAAAGGA---ATAACAATTTAGAGAAAGCAATTTAGTGAANAAT 1397
QY 362 ArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLysCysValGluAsn 381
Db 1398 CAAGAAGAGGTAAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAG 1457
QY 382 ProAsnProThrCysAsnGluAsnAsn-----GlyGlyCysAspAla 395
Db 1458 ATAAAGCCCATTTGTCATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 1517
QY 396 AspAlaThrCysThrGluGluAsp 403
Db 1518 AGTTTGGCATGCACTGAAGATGAA 1541
RESULT 30
US-10-098-841-40
; Sequence 40, Application US/10098841
; Publication NO. US20020197679A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Xu, Chongjun
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunqing
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhao, Qing A.
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-Hong
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhang, Jie
; APPLICANT: Qian, Xiaohong B.
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20020197679A1el Nucleic Acids and
; FILE REFERENCE: 784CIP2
; CURRENT FILING DATE: 2002-03-13
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 331
; SOFTWARE: pc_FL-genes Version 1.0
; SEQ ID NO 40
; LENGTH: 6386
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (150)..(5492)
US-10-098-841-40

Alignment Scores:
Pred. No.: 0.0755 Length: 6386
Score: 129.00 Matches: 81
Percent Similarity: 36.02% Conservative: 53
Best Local Similarity: 21.77% Mismatches: 114
Query Match: 5.67% Indels: 124
DB: 9 Gaps: 16

US-10-057-531A-2 (1-431) x US-10-098-841-40 (1-6386)
```









Qy	63	MetAspAsnIleLeuSerGlyPheGluAnGluTyrAspValIleTyrLeuLysProLeu	82
Db	1778	ATGGATGATTTCCATAAGAAATGTGAAGGGAAGAAAGATTGATCTT	1825
Qy	83	AlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIlePheThrPheAsnLeu	102
Db	1826	-----GAACTGTGGAGAAGAAGTAGAGCAACAATC---CAGTACACACAGT	1867
Qy	103	AsnLeu-----AsnAspIle-----LeuAsnSerArgLeuLysLysArg	115
Db	1868	GACTAGACAAAGGTAAATGAATTACAGAGGAGCTAGAGAGACTTTAAAGAAAG	1927
Qy	116	LysTyrPheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsn	135
Db	1928	GATCAAAATGACCAAAACATAGAAAAACATTATGGTTCAAATCAAAAGTCTCTCTGAACAC	1987
Qy	136	GluTyrIleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeu	155
		::: :::::         :::::	
Db	1988	AAAGAAGTATTGTGAGTGAAGTGAAGTCTCTT---TATGAGGAAACAAATAAACTCAGT	2044
Qy	156	LysSerTyrIleLysGluSerValGlu-----AsnAspIle	169
Db	2045	TCAGAAAAAAGCATTTGAGTAGGAGTATTTGGAGTTTTTTTCTCTCAAAAAGACATTT	2104

```
Qy 170 LysPheAlaGlnGluGlyIleSerTyrTyrGluLysValLeuAlaLysTyrLysAsp 189
Db 2105 ---ATCCTTAAGAACATATCTCAATTAAGAAAGAACTTCAGTTAATGGTTGAAGAC 2161
Qy 190 LeuGluSerIleLysValIleLysGluGlu----- 200
Db 2162 CAGATAATTTAATAAATGCTTGAATAATGAGCAAGTTCAGAACTTATTTGTTAAACT 2221
Qy 201 -----LysGluLysPheProSerSerProProThrThrProProSer 214
Db 2222 CAGTTGTATGTTTCTTAAAGAAATG-----GGATCA 2254
Qy 215 ProAlaLysThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIle 234
Db 2255 GAAGTTTCAGAGACAGTGAAGAGAAAGAT----- 2284
Qy 235 GluThrLeuTyrAsnAsnLeuValAsnLysIleAspTyrLeuIleAsnLeuLysAla 254
Db 2285 -----GTTGTTAATGCTTCTCAGCAGCTCGGTGAATCCTTG-----GCA 2323
Qy 255 LysIleAsnAsp-----CysAsnVal-----GluLysAspGluAlaHisValLysIle 270
Db 2324 AAAATAAATGAGAAATAATGCAACCTGCTTTTCAGCTGATGAAAGAAAGTATTAGAGTTA 2383
Qy 271 ThrLysLeuSerAspLeuLysAlaIleAspAspLysIleAspLeuPheLysAsnProTyr 290
Db 2384 GAAAA-----GAGATTAAAGTGCCTCAAGAA-----GAGAGTGTAGTT 2422
Qy 291 AspPheGluAlaIleLysLysLeuIleAsnAsp---AspThrLysLysAspMetLeuGly 309
Db 2423 CAGTGTGAAGAACTTAAGTCTTTATTGAGACACTATGAGCAAGAAAGTCTCTTAGG 2482
Qy 310 LysLeuLeuSerThrGlyLeuValGlnAsnPheProAsnThrIleIleSerLysLeuIle 329
Db 2483 AAAGAGTTAGAGAA-----ATACAGTCAGAAAAAGAGCCCTGCAGTCTGATCTTCTA 2536
Qy 330 GluGlyLys-----PheGlnAspMetLeu----- 337
Db 2537 GAATGAAGAAAGTCTAATGAAAAACAAGGCTTGAATAATCAGAATCTTTTAAATCAAGTT 2596
Qy 338 ---AsnIleSerGlnHisGlnCysValLysLysGlnCys---ProGluAsnSerGlyCys 355
Db 2597 GAAGAAGTATCTCAA---ACATGTAGCAAAAGTGAATCCATATGAAAGAAAGAAATGT 2653
Qy 356 PheArgHisLeuAspGluArgGluCysLysCysLysLeuLeuAsnTyrLysGlnGluGly 375
Db 2654 TTT-----ATAAGGAACATGAAACCTAAAGCCACTACTAGAACAAAAAGAAATTACGA 2707
Qy 376 AspLysCysValGlu 380
Db 2708 CATAGCAGCAGCAG 2722

RESULT 36
US-09-974-300-2643
; Sequence 2643, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085-500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2643
; LENGTH: 1678
; TYPE: DNA
```

```
; ORGANISM: Bacillus licheniformis
us-09-974-300-2643
```

## Alignment Scores:

```
Pred. No.: 0.0223 Length: 1678
Score: 126.00 Matches: 91
Percent Similarity: 36.32% Conservativeness: 63
Best Local Similarity: 21.46% Mismatches: 128
Query Match: 5.54% Indels: 142
DB: 10 Gaps: 21
```

US-10-057-531A-2 (1-431) x US-09-974-300-2643 (1-1678)

```
Qy 83 AlaGlyValTyrArgSerLeuLysLysGln-----IleGluLysAsnIle 97
Db 277 GCCGACAAATACCGTTTTCAAAAATCGAAACAGGTCGCGCATATTGAGAACTTTTA 336
Qy 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117
Db 337 TCAGCGGCTGAGTCCACATTTGAACATCTTA-----AAGGAA 375
Qy 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137
Db 376 ATGCCGATCTTGTCTCAAGCGAA-----GACGACACCGCAAA 414
Qy 138 IleLeuGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157
Db 415 GAAATCGAAGAGGTAAGAAACGTTATACAAAAGTCGCGAAACCTTTTGGCATACAGC 474
Qy 158 TyrLysTyr-----IleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGly 175
Db 475 CACCTGTACGAGACCTCTATCGCAAAATTTGAAGCCGACCTCGATACCGTGTGGGAAGGC 534
Qy 176 IleSerTyrTyrGlu-----LysValLeu 183
Db 535 ATTAACAGTTTGAAGAGAAACAGACGCGGAATTTACATAGAGCCAGAAAGTGTGTG 594
Qy 184 AlaLysTyrLysAspAspLeuGluSerIleLysLysValIleLysGluGluLysGluLys 203
Db 595 CTTGCCCAAGACCGTTTGTGAAGAGCTTCAATCATACATAGACGATGTGCCGAAACGTG 654
Qy 204 PheProSerSerProProThrThrProProSerProAlaLysThrAspGluGlnLysLys 223
Db 655 CTTGCCAGCTCCAGCAGACAGTTCGCGAGGAAATCGCAAA----- 696
Qy 224 GluSerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrAsnAsnLeuValAsn 243
Db 697 -----TTGAAAGCCGCTATCAGGAAATGATCGAT 726
Qy 244 Lys-----IleAspAspTyrLeuIleAsn 251
Db 727 AAAGGGTATAAGCTCGATCATATTCAAGTGGAGAGGAGCTCGAAACCTTTTGAAGAA 786
Qy 252 LeuLys-----AlaLysIleAsnAspCysAsnValGluLysAspGluAlaHis 267
Db 787 TTGAAGCCGCTGAGATGCGCTTTTGAGCAGCTTGATTAGAGAAAGCCGCCCAATC 846
Qy 268 ValLysIleThrLysLysLeuSerAspLeuLysAlaIleAspAspLysIleAspLeuPheLys 287
Db 847 GTGCAGATC-----ATTGATGAACGATCCAGACGCTGAT 882
Qy 288 AsnProTyrAspPheGlu----- 293
Db 883 AACGAGCTCGAGCATGAAGTCGAAGCAGGCGAGGAGAAATTTTGGGCAAGTGCCTGAGCTT 942
Qy 294 -----AlaIleLysLysLeuIleAsnAspThrLysLysAsp----- 306
Db 943 GCCGCTGCTTGGAAAGCTT-----GAGCCCTCCAAAAAGGATACCCAGACTGAAACC 996
Qy 307 -----MetLeuGlyLysLeuLeuSerThrGlyLeuVal-----GlnAsnPhe 320
Db 997 GAGCTCGTCAAGAAAGGCTACAGACTGACTACCGCGCAACTGGAGAACGACGACGTATAT 1056
```

Qy 321 ProAsnThrIleSerLeuLeuIleGluGlyLysPheGlnAsp----- 335  
Db 1057 GAAAGCGCCTT-----GAAATGATCGAAAGAGAGTTCGAACAGAGTGAAGAAAGGCTT 1110  
Qy 336 -MetLeuAsnIleSerGlnHisGlnCysValLysLysGlnCysPro----- 350  
Db 1111 GATCAGAAACATGTCGCTACTCGTGTGTTAAAGAGAGAGTTCGCGATATCGAAAGAGAG 1170  
Qy 351 -----GluAsnSerGly---CysPheArgHisLeuAspG1 361  
Db 1171 ATGGAAGCGCGCAACGGAGCATGACGAAATACAGGGACATGC----- 1213  
Qy 361 uArgGluCysLysCys-----LeuLeuAsnTyrLysGlnGlu-GlyAspLysCysV 379  
Db 1214 -----TCGAATGCTGAGAAAGAGAGAGCTTCAGGCAAGAGAGCTTCTCAAGCAA 1263  
Qy 379 alGluAsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAlaThrC 399  
Db 1264 TTGAA-----GCAGACAAATCAAGGATACGGCGCGCAGCTCGAGAAAGCAAT 1311  
Qy 399 ysThrGluGluAspSerGlySerArgLysLysIleThrCys-----GluCysThrL 417  
Db 1312 GTTCAGGCATCCCGGAGGCCATAACCGAAAGATCCGACAGTCCGACAGCAGCTGTCGAG 1371  
Qy 417 ysProAsp 419  
Db 1372 AAAGTGAC 1379

## RESULT 37

US-09-880-107-2195  
; Sequence 2195, Application US/09880107  
; Patent NO. US20020142981A1  
; GENERAL INFORMATION:  
; APPLICANT: Horne, Darci T.  
; APPLICANT: Vockley, Joseph G.  
; APPLICANT: Scherf, Uwe  
; APPLICANT: Gene Logic, Inc.  
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer  
; FILE REFERENCE: 44921-5028-WO  
; CURRENT APPLICATION NUMBER: US/09/880,107  
; CURRENT FILING DATE: 2001-06-14  
; PRIOR APPLICATION NUMBER: US 60/211,379  
; PRIOR FILING DATE: 2000-06-14  
; PRIOR APPLICATION NUMBER: US 60/237,054  
; PRIOR FILING DATE: 2000-10-02  
; NUMBER OF SEQ ID NOS: 3950  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2195  
; LENGTH: 7573  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 L33075  
US-09-880-107-2195

Alignment Scores:  
Pred. No.: 0.202 Length: 7573  
Score: 125.50 Matches: 69  
Percent Similarity: 37.97% Conservative: 81  
Best Local Similarity: 17.47% Mismatches: 106  
Query Match: 5.52% Indels: 139  
Gaps: 16

US-10-057-531a-2 (1-431) x US-09-880-107-2195 (1-7573)

Qy 75 AspValIleTyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGlu 94  
Db 2889 GAAGTGTAAAGATTTCAGTCCCTGGCAAGAGTGCACCAAGCT---CGAAAGCGCTATCGA 2945  
Qy 95 LysAsnIlePheThrPheAsnLeuAsnLeuAsnGlyLysPheGlnAsp----- 112  
Db 2946 GATCCCTGCAGTACTCCGGGACCATATAATGACATTATCAAAATCCAGCTTTTATT 3005

Qy 113 -----LysLysArgLysTyrPheLeuAspValLeuGluSerAsp----- 125  
Db 3006 CGGCAAAACAAAGCTCGGAGTACAGAGCTCTCAATCAATGCTGAGAGCTCCTCTATG 3065  
Qy 126 -----LeuMetGlnPheLysHisIle-----SerSerAsnGluTyrIleIleGluAsp 141  
Db 3066 GTTGTGTCGCAAAATTTGTCACCTGCTGGACCAAAAGTGCACAGGATTTTCAGAGGAG 3125  
Qy 142 SerPheLysLeuLeuAsnSerGlnGlnLysAsnThrLeuLeuLysSerTyrLysTyrIle 161  
Db 3126 CTTGACCTTATGAAGATCGGGAAGAGGTTATCACCCCTCATCTCTCT----- 3173  
Qy 162 LysGluSerValGluAsnAspIleLysPhe----- 171  
Db 3174 AACAGCAGCTGGAGATGACCTCAATCTCATGGATATCAAAATTTGGACTGCTAGTGA 3233  
Qy 172 -----AlaGlnGluGlyIleSerTyrTyrGluLysValLeuAlaLysTyrLys 187  
Db 3234 AATAAGATTACGTTGTCAGGATGTGTTTCCACAGTAAAAAACTTACCAAAAAAATAAG 3293  
Qy 188 AspAspLeuGlu-----SerIleLysLysValIle 197  
Db 3294 GAACAGTTGCTGATATGATGATAATAAACAAGAGGAGTCTCAAGCTTTTGAGC 3353  
Qy 198 LysGluLysLysGluLysPheProSerSerPro-----ProThrThr 211  
Db 3354 AAGAGAAAGAGAGAGAGTTCGAAGCTTACCAGCAGCTGTTTATTATTATGCAACCAAT 3413  
Qy 212 ProProSerProAlaLysThrAspGluGlnLysLysGluSerLysPheLeuProPheLeu 231  
Db 3414 CCCACCTATCTGGCCAAAGCTCATTTTTCAGATGCCCCAGAACAAAGTCCACCAGTTCTATG 3473  
Qy 232 ThrAsn---IleGluThrLeuTyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIle 250  
Db 3474 GACTCTGTAACTTTCACACTCTACAACTACGGCTCCACACCGAGAGAGTACTGCTC 3533  
Qy 251 -----AsnLeuLysAlaLysIleAsnAspCysAsn 260  
Db 3534 CTGCGGCTCTTTAAGACAGCAGCTCCCAAGAGGAAATCAAGTCGAAGGTAGATCAGATTCAA 3593  
Qy 261 -----ValGluLysAspGluAlaHisValLys----- 269  
Db 3594 GAGATTGTGACAGAGAAATCCTACGGTTATTAAATGGTTTGAAGTTTCAACCGTGGTGC 3653  
Qy 270 -----IleThrLysSerAspLeu 276  
Db 3654 CGTGCCCAAGTGCCTGAGACAGATCTTGCCCCCAGCTCGTGAAGAAATTTATGATGAC 3713  
Qy 277 LysAlaIleAsp-----AspLysIleAspLeuPheLysAsn----- 288  
Db 3714 AAATCTCTCAACATCAAAATGACCTGTGGATATTACAAATCTTGGGTTAATCAGATG 3773  
Qy 289 -----ProTyrAspPheGluAlaIleLysLysLeu 298  
Db 3774 GAGTCTCAGACAGGAGGCAAGCAAACTGCCCTATGATGTGACCCCTGAGCAGCGCTA 3833  
Qy 299 IleAsnAspAspThrLysLysAspMet----- 307  
Db 3834 GCTCATGAAGAAGTCAACAGACAGCGGTAGACAGCTCCATCAGCAACATCGGGCTGTGACA 3893  
Qy 307 ----- 307  
Db 3894 GACAAATTTCTCTCAGCCATTGTACGCTCTGTGGACAAAATCCCTTATGGGATGCGCTTC 3953  
Qy 308 LeuGlyLysLeuLeuSerThrGlyLeuValGlnAsnPheProAsnThr----- 323  
Db 3954 ATTGCCAAAGTGTCAAGAGACTGTTGTCATGAGAAGTTCCCTGCTGGTGGATGAG 4013  
Qy 324 -----IleIleSerLysLysIleGluGlyLysPheGlnAsp 335  
Db 4014 CTGCTGAAGATTATTGGTAACTTGTCTTTATTATCATCATCAAT 4058

RESULT 38

US-10-098-514-16  
; Sequence 16, Application US/10098514  
; Publication No. US20020194648A1  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Sandra P  
; APPLICANT: Christopher, David A  
; APPLICANT: Vine, Benjamin  
; APPLICANT: Su, Wei-wen  
; APPLICANT: Bugos, Robert  
; TITLE OF INVENTION: PLASMODIUM FALCIPARUM MEROZOITE SURFACE PROTEIN-1 MALARIA PRODUCE  
; FILE REFERENCE: A-71339/RET/TAL/NBC  
; CURRENT APPLICATION NUMBER: US/10/098,514  
; PRIOR FILING DATE: 2002-08-06  
; PRIOR APPLICATION NUMBER: US 09/500,376  
; PRIOR FILING DATE: 2000-02-08  
; PRIOR APPLICATION NUMBER: US 60/274,599  
; PRIOR FILING DATE: 2001-03-09  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 16  
; LENGTH: 142  
; TYPE: DNA  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: synthetic  
US-10-098-514-16

Alignment Scores:  
Pred. No.: 0.000893 Length: 142  
Score: 125.00 Matches: 21  
Percent Similarity: 76.09% Conservative: 14  
Best Local Similarity: 45.65% Mismatches: 11  
Query Match: 5.49% Indels: 0  
DB: 9 Gaps: 0

US-10-057-531A-2 (1-431) x US-10-098-514-16 (1-142)

QY 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137  
Db 3 TTTCAAAATGTTTAAAGATCAGATTAAATTCATATAAAGATTAAACATCAAGTAATTAT 62  
QY 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157  
Db 63 GTTGTCAAAGATCCATATAAATTTCTTAATAAGAAAGAGATAAATTCCTTAAGCAGT 122  
QY 158 TyrLysTyrIleLysGlu 163  
Db 123 TATAATTATTAAGGAT 140

RESULT 39  
US-09-298-523B-70  
; Sequence 70, Application US/09298523B  
; Publication No. US20030059438A1  
; GENERAL INFORMATION:  
; APPLICANT: BRILES et al.  
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS  
; FILE REFERENCE: 454312-3140  
; CURRENT APPLICATION NUMBER: US/09/298,523B  
; CURRENT FILING DATE: 1999-04-23  
; NUMBER OF SEQ ID NOS: 78  
; SOFTWARE: PatentIn ver. 2.1  
; SEQ ID NO 70  
; LENGTH: 2405  
; TYPE: DNA  
; ORGANISM: Streptococcus pneumoniae  
US-09-298-523B-70

Alignment Scores:  
Pred. No.: 0.0698 Length: 2405  
Score: 123.00 Matches: 81  
Percent Similarity: 38.30% Conservative: 45

Best Local Similarity: 24.62% Mismatches: 147  
Query Match: 5.41% Indels: 56  
DB: 9 Gaps: 12  
US-10-057-531A-2 (1-431) x US-09-298-523B-70 (1-2405)  
QY 19 LysGluThrAlaAlaLysPheGluArgGlnHisMetAspSerProAspLeuGlyThr 38  
Db 973 AAAGAGCAGGTAACCAAGAGCGGGCAACAGGAGGAGTTCTCGACAGCTAGCAACA 1032  
QY 39 AspAsp-----AspAspLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThr 56  
Db 1033 CCTGATAAAAAAGAAATGATCGCAAGTCTTCAGATTCTGAGGTAGGTGAAGAACTCTT 1092  
QY 57 MetAlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspVal 76  
Db 1093 CCAAGCCCATCCCTTAATATGCAAT---GAAAGTCACACAGACATAGGAAGATGTC 1149  
QY 77 ---IleTyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLys 95  
Db 1150 GATGAATATATAAAAAAATGTTGAGT-----GAGATCCAATTAGATAGA 1194  
QY 96 AsnIlePheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArg 115  
Db 1195 AGAAAAACATACCCAAATGTCACTTTAAACATAAAGTTGAGCGCAATTAACAGGAAGTAT 1254  
QY 116 LysTyrPheLeuAspValLeuGluSerAsp----- 125  
Db 1255 TTGTATGAATTAAAGTCTTTTAAAGAGAGAACTCGAAAAAGAGAGTTGACGTCAAAAACC 1314  
QY 126 -----LeuMetGlnPheLys 130  
Db 1315 AAAGCAGAGTTAAACCGCAGCTTTTTCAGCAGTTTAAAAAAAGATACATTGAACCCAGAAAAA 1374  
QY 131 HisIleSerSerAsnGluTyrIleIleGluAspSerPheLysLeuLeu----- 146  
Db 1375 AAGTAGCAGCAAGCTGAGAGAGAGTTGAAGAGAGTTGAAGAGAGAGAGAGAGAGAGAGAGAGAG 1434  
QY 147 AsnSerGluGlnLysAsnThrLeuLeuLysSerTyrLysTyrIleLys---GluSerVal 165  
Db 1435 GAAGAAGATCGCGTAACACCAACCAATCTTACAAAAACGCTTCAACTTGAATTTGCT 1494  
QY 166 GluAsnAspIleGlyPheAlaGlnGluGlyIleSerTyrTyrGluLysValLeuAlaLys 185  
Db 1495 GAGTCCGATGTGAAA-----GTTAAAAAAGCGAGAGCTTGAACCTAGTAAAAA 1539  
QY 186 TyrLysAspLeuGluSerIleLysLysValIleLysGluGluLysGluLysPhePro 205  
Db 1540 GAGGAAGCTAAGCAATCTCGAAACGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1599  
QY 206 Ser---SerProThrThrProProSerProAlaLysThrAspGluGlnLysLysGlu 224  
Db 1600 ACTAAAAAGCTGAGGCTACAAAGTTAGAAAAAATCAACACAGATCGTAAAAAGCAGAA 1659  
QY 225 SerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrAsnAsnLeuValAsnLys 244  
Db 1660 GAAGAAGCTAAACGAAAAAGCAGAGAAATCTGAGAAAAAAGCTGCTGAAGCCCAACAAAAA 1719  
QY 245 IleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsnValGluLysAsp 264  
Db 1720 GTGGATGCTGAAGATATATCTCTTGAAGCTAAATCGTGAAGTTGGAATATGAGTTTCAG 1779  
QY 265 GluAlaHisValLysIleThrLysLeuSerAspLeuLysAlaIleAsp-----AspLys 282  
Db 1780 AGACTAGAA-----AAAGACTCAAGAGAGATTGATCAGTCTGACTCA 1821  
QY 283 IleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLysLeuIleAsnAspAsp 302  
Db 1822 GAAGATTTATTTAAAGAGAGCGCTCGTGTCTCTTCAATCTAAATTTG-----GAT 1872  
QY 303 ThrLysLysAspMetLeuGlyLysLeu 311  
Db 1873 ACCAAAAAGCTAAACTATCAAAACTT 1899

## RESULT 40

US-09-298-523B-59  
; Sequence 59, Application US/09298523B  
; Publication No. US20030059438A1  
; GENERAL INFORMATION:  
; APPLICANT: BRILES et al.  
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS  
; TITLE OF INVENTION: AND STRAINS THEREOF AND USES THEREFOR  
; FILE REFERENCE: 454312-3140  
; CURRENT APPLICATION NUMBER: US/09/298,523B  
; CURRENT FILING DATE: 1999-04-23  
; NUMBER OF SEQ ID NOS: 78  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 59  
; LENGTH: 3463  
; TYPE: DNA  
; ORGANISM: Streptococcus pneumoniae  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (319)..(3105)  
; OTHER INFORMATION: coding sequence for PspA  
US-09-298-523B-59

## Alignment Scores:

Pred. No.:	0.116	Length:	3463
Score:	123.00	Matches:	81
Percent Similarity:	37.06%	Conservative:	45
Best Local Similarity:	23.8%	Mismatches:	148
Query Match:	5.41%	Indels:	66
DB:	9	Gaps:	13

US-10-057-531A-2 (1-431) x US-09-298-523B-59 (1-3463)

QY	18	MetLysGluThrAlaAlaLysPheGluArgGlnHisMetAspSerPro-----	34
DB	1051	TTGAAGGAAGCTGTTGAAAAAGATGTAGCGACTTCAGACCAAGATAAACCAAGAGCGCG	1110
QY	35	-----AspLeuGlyThrAspAsp-----AspAspLysAlaMet	45
DB	1111	GCANAAACGAGGAGTTCTCGAGAGCTAGCAACACCTGATAAAGAAAGAAATATGCGCAAG	1170
QY	46	AlaAspIleGlySerIleGluGlyArgGlyThrMetAlaIleSerValThrMetAspAsn	65
DB	1171	TCTTCAGATTCTAGCGTAGGTGAAGAAACTCTTCCAAGCCCATCCCTTAATATGCGCAAT	1230
QY	66	IleLeuSerGlyPheGluAsnGluTyAspVal---IleTyLeuLysProLeuAlaGly	84
DB	1231	---GAAAGTCACACACACATAGGAAGATGTCGATGAATATATAAAAAAATGTTGAGT	1287
QY	85	ValTyArgSerLeuLysLysGlnIleGluLysAsnIlePheThrPheAsnLeuLeu	104
DB	1288	-----GAGATCCAAATAGATAGAAGAAACATACCCAAATGTCAACTTA	1332
QY	105	AsnAspIleLeuAsnSerArgLeuLysLysArgLysTyPheLeuAspValLeuGluSer	124
DB	1333	AACATAAACTTGCGCCAATTAAACGCAAGTATTCTATGAATTAAGTCTTTTAAAGAG	1392
QY	125	Asp-----	125
DB	1393	AACTCGAAAAAAGAGAGTTGACGTCAAAACCAAGACAGGTTAACCGCAGCTTTTGAG	1452
QY	126	-----LeuMetGlnPheLysHisIleSerSerAsnGluTyIleIle	139
DB	1453	CAGTTTAAAAAGATACATTGAACCCAGAAAAAGGTAGACGAGCTGAGAAGAGGTT	1512
QY	140	GluAspSerPheLysLeuLeu-----AsnSerGluGlnLysAsnThrLeuLeu	155
DB	1513	GAAGAAGCTAAGAAAAAAGCCAGGATCAAAAGAAAGAGATCGCGTAACTACCCCAACC	1572
QY	156	LysSerTyrlsTyrlsIleLys---GluSerValGluAsnAspIleLysPheAlaGlnGlu	174
DB	1573	AATACTTACAAAAACGCTTGAACCTTGAATTTGCTGAGTCGGATGTGAAA-----	1620

Search completed: May 19, 2003, 17:01:06

Job time : 189 secs

QY	175	GlyIleSerTyTyTyGluLysValLeuAlaLysTyTyLysAspLeuGluSerIleLys	194
DB	1621	---GTTAAAGACCGGAGCTTGAACCTAGTAAAGAGAGAGCTAACCAATCTCGAAACGAG	1677
QY	195	LysValIleLysGluGluLysGluLysPheProSer---SerProProThrThrProPro	213
DB	1678	GAAGAAATTAAGCAACGACAAAGAGAAAGTTGAGAGTAAAAAGCTCAGGCTACAAGTTA	1737
QY	214	SerProAlaLysThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsn	233
DB	1738	GAAGAAATCAAGACAGATCGTAAAAAAGCAGAGAAAGCTAAACGAAAAAGCAGAGAA	1797
QY	234	IleGluThrLeuTyTyAsnAsnLeuValAsnLysIleAspAspTyTyLeuIleAsnLeuLys	253
DB	1798	TCTGAGAAAAAGCTCTGCTGAAGCCAAACAAAGTGGATGCTGCAAGAAATATGCTCTTGA	1857
QY	254	AlaLysIleAsnAspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeu	273
DB	1858	GCTAAATCGCTGAGTTGGAATATGAAGTTTCAGAGACTAGAA-----	1899
QY	274	SerAspLeuLysAlaIleAsp-----AspLysIleAspLeuPheLysAsnProTyTyAsp	291
DB	1900	AAAGAGCTCAAGACAGATTGATGAGTCTGACTCAGAAAGATTATCTTAAAGAACGCTCCGT	1959
QY	292	PheGluAlaIleLysLysLeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeu	311
DB	1960	GCTCCTCTTCAATCTAAATTG-----GATACCAAAAAAGCTAAACTATCAAACTT	2010



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 19, 2003, 15:02:45 ; Search time 1277 Seconds  
(without alignments)  
5466.135 Million cell updates/sec

Title: US-10-057-531A-2  
Perfect score: 2275  
Sequence: 1 MHMHHSGLVPRGSMKE.....TCECTKPDSPFLDFGIFCSS 431

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 16134066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlh  
-O/cnp2\_1/USPTO.spool/US10057531/runat\_12052003\_091031\_21676/app\_query.fasta\_1.583  
-DB=EST -QPM=fastap -SUFFIX=est -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=40 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10057531.ecgn\_1.1456 @runat\_12052003\_091031\_21676 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	848	37.3	537	13	BI815756
2	736	32.4	419	14	BQ451367
3	618	27.2	402	13	BI670649
4	555	24.4	365	14	N97742
5	484	21.3	432	14	BQ633144
6	408	17.9	501	13	BI814631
7	403	17.7	422	14	BQ451709
8	398	17.5	522	14	BQ452445
9	393	17.3	455	14	BQ451036
10	317	13.9	303	13	BI815239
11	310.5	13.6	778	13	BI8164106
12	281	12.4	408	13	BI814931
13	281	12.4	418	14	BQ451572
14	281	12.4	422	14	BQ451661
15	279	12.3	595	13	BI677799
16	272	12.0	527	13	BI603032
17	272	12.0	784	13	BI66681
18	249.5	11.0	525	13	BI65325
19	243.5	10.7	439	13	BI60668
20	238.5	10.5	648	13	BI094652
21	238.5	10.5	661	13	BI094593
22	238.5	10.5	686	13	BI094577
23	234.5	10.3	415	13	BI094637
24	234.5	10.3	660	12	BF298872
25	233.5	10.3	405	13	BI094651
26	233	10.2	725	13	BI60032
27	229.5	10.1	542	13	BI67382
28	228.5	10.0	629	13	BI094603
29	225.5	9.9	571	13	BI094610
30	223	9.8	619	13	BI63150
31	203	8.9	782	13	BI64097
32	195	8.6	329	13	BI67649
33	192.5	8.5	624	13	BI094571
34	190.5	8.4	780	13	BI68979
35	189.5	8.3	693	13	BI094596
36	188	8.3	608	13	BI094595
37	184	8.1	652	13	BI66573
38	176.5	7.8	637	13	BI094585
39	166	7.3	362	14	BQ451205
40	164	7.2	686	13	BI67595
41	161	7.1	352	14	BQ577302
42	144.5	6.4	1067	17	CNS05PAD
43	144	6.3	558	13	BI094573
44	144	6.3	613	13	BI094586
45	133.5	5.9	699	9	AU061475

ALIGNMENTS

RESULT 1  
BI815756  
LOCUS  
DEFINITION pfEST00a32c09.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium falciparum cDNA 5' similar to TR:Q9TVG8 Q9TVG8 MSA1 PROTEIN ;, mRNA sequence.  
BI815756 537 bp mRNA linear EST 03-OCT-2001  
BI815756.1 GI:15909334  
EST.  
malaria parasite P. falciparum.  
Plasmodium falciparum  
ORGANISM Plasmodium falciparum  
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
REFERENCE 1 (bases 1 to 537)  
AUTHORS Tang.K., Cole.R., Chakrabarti.D., Haywood.R., Clifton.S., Pape.D., Marra.M., Hillier.L., Martin.J., Wylie.T., Dante.M., Theising.B.,

Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Jentes, E., Ronko, I., Tsagareishvili, R., Belaygorod, L., Franklin, C., Carr, L., Grow, A., Maguire, L., Richey, J., Wadkins, J., Kennedy, S., Levinso, D., Waterston, R., Wilson, R. and Sibley, D.

WashU Plasmodium EST Project  
Unpublished (2001)  
Contact: L. David Sibley  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

Library was constructed by Debopam Chakrabarti DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: L. David Sibley (sibley@borcim.wustl.edu), Washington University  
Seq primer: -40UP from Gibco  
High quality sequence stop: 422.

## FEATURES

source

1. .537  
Location/Qualifiers  
/organism="Plasmodium falciparum"  
/db\_xref="taxon:5833"  
/clone\_lib="Plasmodium falciparum 3D7 asexual cDNA"  
/lab\_host="DH10B (GeneHog, Invitrogen, Inc.)"  
/note="Vector: pBluescript SK plus; Site\_1: EcoRI; Site\_2: XhoI; Library was constructed by Debopam Chakrabarti.  
Total RNA samples were isolated from mixed stage saponin(0.1%) -lysed P. falciparum 3D7 infected erythrocytes by the acidic guanidinium-phenol chloroform method. The poly A+ RNA was isolated by the polyAT-Tract mRNA isolation system (Promega, WI) using streptavidin Magnesphere particles. Directional cDNA libraries were constructed by oligo d(T) priming of poly(A)+ RNA (5mg) into EcoRI and XhoI sites of 1 ZapII vector using the Zap cDNA synthesis kit (Stratagene, CA). The average size of the cDNA inserts in the library was between 1.0 and 1.5kb. Clones were mass excised using the ExAssist helper phage (Stratagene), the phagemids were precipitated with PEG 8000 and extracted with phenol/chloroform. Phagemid DNA was electroporated into DH10B cells."

BASE COUNT 217 a 75 c 83 g 162 t

## Alignment Scores:

Pred. No.: 4.84e-83 Length: 537  
Score: 848.00 Matches: 154  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.27% Indels: 0  
DB: 13 Gaps: 0

US-10-057-531a-2 (1-431) x B1815756 (1-537)

Qy 278 AlaileAspAspLysileAspLeuPheLysAsnProTyrAspPheGluAlaileLysLys 297  
Db 2 GCATTGATGACAAATAGACTTTTAAACCCCTTACGACTTCGACCAATTAATAAAA 61  
Qy 298 LeuileAsnAspThrlsLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal 317  
Db 62 TTGATAATGATGATACGAAACAAATATGCTTGCCAAATTTACTAGTACAGATTAGTT 121  
Qy 318 GlnAsnPheProAsnThrlleleSerLysLeuileGluGlyLysPheGlnAspMetLeu 337  
Db 122 CAAAATTTTCCTAAACAAATATATCAAAATTAATTAAGGAAAAATTCACAGATATGTTA 181  
Qy 338 AsnileSerGlnHisGlnCysVallyLysGlnCysProGluAsnSerGlyCysPheArg 357  
Db 182 ACATTTTCCACACACCAATGCGTAAACAAATGCTCCAGAAATTCGAGATTTCAGA 241  
Qy 358 HisLeuAspGluArgGluGlyCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377  
Db 242 CATTAGATGAAGAGAGAGATGTAATGTTTATTAATTAACAACAGAGGTGATAAA 301

Qy 378 CysValGluAsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaaspAla 397  
Db 302 TGTGTTGAATCCAAATCCTACTTGTAAACGAAATAATGGTGGATGTGATGCAGATGCC 361  
Qy 398 ThrCysThrGluGluAspSerGlySerArgLysLysileThrCysGluCysThrLys 417  
Db 362 ACATGTACCGAAGAAGATTTCAGGTAGCAGCAGAGAAATACATGTGAATGTACTAAA 421  
Qy 418 ProAspSerTyrProLeuPheAspGlyLysilePheCysSerSer 431  
Db 422 CCTGATTCTTATCCACTTTTCGATGTGATTTTCTGCAGTTCC 463

## RESULT 2

BO451367 419 bp mRNA linear EST 29-MAY-2002  
LOCUS PFEST0aa98c03.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium  
DEFINITION falciparum cDNA 5' similar to TR:Q9TVG8 Q9TVG8 MSA1 PROTEIN ; mRNA  
sequence.  
BO451367  
VERSION BO451367.1 GI:21254479  
KEYWORDS EST.  
SOURCE malaria parasite P. falciparum.  
ORGANISM Plasmodium falciparum  
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
REFERENCE 1 (bases 1 to 419)  
AUTHORS Tang, K., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Pape, D., Marra, M., Hillier, L., Martin, J., Wylie, T., Dante, M., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Jentes, E., Ronko, I., Tsagareishvili, R., Belaygorod, L., Franklin, C., Carr, L., Grow, A., Maguire, L., Richey, J., Wadkins, J., Kennedy, S., Levinso, D., Waterston, R., Wilson, R. and Sibley, D.

WashU Plasmodium EST Project  
Unpublished (2001)  
Contact: L. David Sibley  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Library was constructed by Debopam Chakrabarti DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: L. David Sibley (sibley@borcim.wustl.edu), Washington University  
Seq primer: -40UP from Gibco.

## TITLE

## JOURNAL

## COMMENT

## FEATURES

source

1. .419  
Location/Qualifiers  
/organism="Plasmodium falciparum"  
/db\_xref="taxon:5833"  
/clone\_lib="Plasmodium falciparum 3D7 asexual cDNA"  
/lab\_host="DH10B (GeneHog, Invitrogen, Inc.)"  
/note="Vector: pBluescript SK plus; Site\_1: EcoRI; Site\_2: XhoI; Library was constructed by Debopam Chakrabarti.  
Total RNA samples were isolated from mixed stage saponin(0.1%) -lysed P. falciparum 3D7 infected erythrocytes by the acidic guanidinium-phenol chloroform method. The poly A+ RNA was isolated by the polyAT-Tract mRNA isolation system (Promega, WI) using streptavidin Magnesphere particles. Directional cDNA libraries were constructed by oligo d(T) priming of poly(A)+ RNA (5mg) into EcoRI and XhoI sites of 1 ZapII vector using the Zap cDNA synthesis kit (Stratagene, CA). The average size of the cDNA inserts in the library was between 1.0 and 1.5kb. Clones were mass excised using the ExAssist helper phage (Stratagene), the phagemids were precipitated with PEG 8000 and extracted with phenol/chloroform. Phagemid DNA was electroporated into DH10B cells."

BASE COUNT 189 a 49 c 59 g 122 t

## ORIGIN

Alignment Scores: 7.4e-71 Length: 419  
Pred. No.: 7.4e-71



Score: 736.00 Matches: 139  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 32.35% Indels: 0  
 DB: 14 Gaps: 0

US-10-057-531A-2 (1-431) x B0451367 (1-419)

QY 237 LeuTyrAsnAsnLeuValAsnLysIleAspTyrLeuIleAsnLeuLysAlaLysIle 256  
 DB 2 TTATACAAATAGTACTTAATAAATGACGATTACTTAATAAATGACGAAAGATT 61  
 QY 257 AsnAspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeu 276  
 DB 62 AACGATTGTAATGTTGAAAGATGACACATGTTAAATAAATACTAACTTACTGATTGA 121  
 QY 277 LysAlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaLys 296  
 DB 122 AAAGCAATTTGATGACAAATAGATCTTTTAAACACCTTACGACTTCCGAAGCAATTA 181  
 QY 297 LysLeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuSerThrGlyLeu 316  
 DB 182 AAATTGATAAATGATGATACGAAAGATATGCTTGCGCAATTAATTACTTAGTACAGATT 241  
 QY 317 ValGlnAsnPheProAsnThrIleIleSerLysLeuIleGluLysPheGlnAspMet 336  
 DB 242 GTTCAAAATTTTCTTAATCAATAATATCAAAATTAATTAAGGAAATTTCCAAGATATG 301  
 QY 337 LeuAsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPhe 356  
 DB 302 TTAACATTTTCAACACCAATGCGTAAACAAACATGTCGAGAAATTTCTGGATGTTTC 361  
 QY 357 ArgHisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGly 375  
 DB 362 AGACATTTAGATGAAGACAGAAATGTAATGTTTATTAAATTAACAAAGAGT 418

RESULT 3  
 B1670649 402 bp mRNA linear EST 12-SEP-2001  
 LOCUS pfEST00a02ell.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium  
 DEFINITION falciparum cDNA 5' similar to TR:043995 043995 MERZOITE SURFACE  
 PROTEIN-1 ; mRNA sequence.

ACCESSION B1670649  
 VERSION B1670649.1 GI:15586033  
 KEYWORDS EST.  
 SOURCE malaria parasite P. falciparum.  
 ORGANISM Plasmodium falciparum  
 REFERENCE 1 (bases 1 to 402)  
 AUTHORS Tang, K., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Pape, D.,  
 Marra, M., Hillier, L., Martin, J., Wylie, T., Dante, M., Theising, B.,  
 Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Jentes, E., Ronko, I.,  
 Tsagaris, V., Richey, J., Belyakov, L., Franklin, C., Carr, L., Grow, A.,  
 Maguire, L., Richey, J., Wadkins, J., Kennedy, S., Levinso, D.,  
 Waterson, R., Wilson, R. and Sibley, D.  
 TITLE WashU Plasmodium EST Project  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: L. David Sibley  
 WashU Plasmodium EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.wustl.edu

Library was constructed by Debopam Chakrabarti DNA sequencing by:  
 Washington University Genome Sequencing Center For information on  
 obtaining a clone please contact: L. David Sibley  
 (sibley@borcim.wustl.edu), Washington University  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 352.  
 Location/Qualifiers  
 1..402  
 /organism="Plasmodium falciparum"

FEATURES  
 source

/db\_xref="taxon:5833"  
 /clone\_lib="Plasmodium falciparum 3D7 asexual cDNA"  
 /lab\_host="DH10B (GeneHog, Invitrogen, Inc.)"  
 /note="Vector: pBluescript SK plus; Site\_1: EcoRI; Site\_2:  
 XhoI; Library was constructed by Debopam Chakrabarti.  
 Total RNA samples were isolated from mixed stage  
 saponin(0.1%)-lysed P. falciparum 3D7 infected  
 erythrocytes by the acidic guanidinium-phenol chloroform  
 method. The poly A+ RNA was isolated by the polyAT-Tract  
 mRNA isolation system (Promega, WI) using streptavidin  
 MagSphere particles. Directional cDNA libraries were  
 constructed by oligo d(T) priming of poly(A)+ RNA (5mg)  
 into EcoRI and XhoI sites of 1 zapII vector using the Zap  
 cDNA synthesis kit (Stratagene, CA). The average size of the  
 cDNA inserts in the library was between 1.0 and 1.5kb.  
 Clones were mass excised using the Exassist helper phase  
 (Stratagene), the phagemids were precipitated with PEG  
 8000 and extracted with phenol/chloroform. Phagemid DNA  
 was electroporated into DH10B cells."

BASE COUNT 186 a 56 c 44 g 111 t 5 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 6.87e-58 Length: 402  
 Score: 618.00 Matches: 124  
 Percent Similarity: 94.03% Conservative: 2  
 Best Local Similarity: 92.54% Mismatches: 4  
 Query Match: 27.16% Indels: 4  
 DB: 13 Gaps: 1

US-10-057-531A-2 (1-431) x B1670649 (1-402)

QY 97 IlePheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLys 116  
 DB 1 ATTTTACATTTTAAATTTTAAATTTTGAACGATATCTTAAATTCAGCTTTAAGAACGAA 60  
 QY 117 TyrPheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGlu 136  
 DB 61 TATTTCTTAGATCTATTAGCAATCTGATTAAATGCAATTTAAACATATATCCCAATGAA 120  
 QY 137 TyrIleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLys 156  
 DB 121 TACATTTATTGAAGATTCATTTAAATTTATTGAATTCAGAACAAAAACACACTTTTAAA 180  
 QY 157 SerTyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIle 176  
 DB 181 AGTTACAAATATATAAAGAAATCAGTAGAAATGATATTAAATTTCCAGGAGGTATA 240  
 QY 177 SerTyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysVal 196  
 DB 241 AGTTATTATGAAAGGTTTTAGCGAAATATAAGGATGATTAGATCAATTAATAAAGTT 300  
 QY 197 IleLysGluLysGluLysPheProSerSerProThrThrProProSerProAla 216  
 DB 301 ATCAAGAGAGAAAGGAGAGATTCCTCATCATCACCACCAACACACCTCCGCACGAC 360  
 QY 217 LysThrAsp-----GluGlnLysLysGluSerLys 226  
 DB 361 NNNNAACCAACCCCAACATCCCAACAAAAAAGAAAAA 402

RESULT 4  
 N97742 365 bp mRNA linear EST 18-NOV-1996  
 LOCUS N97742  
 DEFINITION 1183C3 czappFDd2.1, Debopam Chakrabarti Plasmodium falciparum cDNA  
 clone PF1183C, mRNA sequence.

ACCESSION N97742  
 VERSION N97742.1 GI:1674760  
 KEYWORDS EST.  
 SOURCE malaria parasite P. falciparum.  
 ORGANISM Plasmodium falciparum  
 REFERENCE 1 (bases 1 to 365)  
 AUTHORS Dame, J.B., Arnot, D.E., Bourke, P., Chakrabarti, D., Christodoulou, Z.,

Coppel, R., Cowman, A., Craig, A., Fischer, K., Foster, J., Goodman, N., Hinterberg, K., Holder, A.A., Holt, D., Kemp, D., Lanzer, M., Lim, A., Newbold, C., Ravetch, J.V., Reddy, G.R., Rubio, J., Schuster, S.M., Su, X.-Z., Thompson, J.K., Vital, F., Wellens, T.E., and Werner, E.  
Current status of the Plasmodium falciparum genome project  
Mol. Biochem. Parasitol. 79, 1-12 (1996)  
97001675

COMMENT  
Contact: Debopam Chakrabarti  
Department of Molecular Biology and Microbiology  
University of Central Florida  
Orlando, FL 32816-2360  
Tel: 407 384 2061  
Fax: 407 384 3095  
Email: dcchak@pegasus.cc.ucf.edu  
Seq primer: T3.

FEATURES  
source  
1. .365  
/organism="Plasmodium falciparum"  
/strain="Dd2"  
/db\_xref="taxon:5833"  
/clone="PF1183C"  
/lab\_host="E. coli XL-1 blue"  
/note="Vector: Lambda ZAP II; Site\_1: EcoR I; Site\_2: Xho I; PolyA+ RNA, from asynchronous blood stage parasites of the Dd2 isolate cultured in vitro, was reverse transcribed using an oligo dT-Xho I primer. Second strand was prepared using RNase H and DNA polymerase I. EcoR I adapters were ligated to the cDNA, and it was digested with Xho I. Prepared fragments were ligated into EcoR I + Xho I digested lambda ZAP II vector."  
BASE COUNT 163 a 39 c 43 g 120 t

Alignment Scores:  
Pred. No.: 5.2e-51 Length: 365  
Score: 555.00 Matches: 111  
Percent Similarity: 99.11% Conservations: 0  
Best Local Similarity: 99.11% Mismatches: 1  
Query Match: 24.40% Indels: 0  
DB: 14 Gaps: 0

US-10-057-531A-2 (1-431) x N97742 (1-365)

Qy 58 AlarlservValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyAspValile 77  
|||||  
Db 28 GCAATATCTGTCACATGGATAATATCTCTCAGGATTTGAAATGAATGATGTTATA 87  
|||||  
Qy 78 TyrLeuLysProLeuAlaGlyValTyArgSerLeuLysLysGlnIleGluLysAsnIle 97  
|||||  
Db 88 TATTTAAACCTTTAGCTGGAGTATATAGAAGCTTAAACAAATTTGAAAAAACATT 147  
|||||  
Qy 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117  
|||||  
Db 148 ATTACATTTAATTAATTTGACGATATCTTAATTCACGCTTTAAGAACGAAATAT 207  
|||||  
Qy 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137  
|||||  
Db 208 TTCCTTAGATGATTAGAATCTGATTTAATGCAATTTAAACATATATCCTCAATGAATAC 267  
|||||  
Qy 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157  
|||||  
Db 268 ATTTATGAAGATTCATTTAATTTAATTCAGAACAAAAACACACTTTTAAAAAGT 327  
|||||  
Qy 158 TyrLysTrileLysGluSerValGluAsnAspIle 169  
|||||  
Db 328 TACAATATATAAAGAATCATAGTAAAAATGATATT 363  
|||||

RESULT 5  
BQ633144 432 bp mRNA linear EST 03-JUL-2002  
LOCUS PFESToab38c11.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium  
DEFINITION falciparum cDNA 5' similar to TR:Q9UA18 Q9UA18 SURFACE PROTEIN-1 ;

ACCESSION  
BQ633144  
VERSION  
BQ633144.1 GI:21687297  
KEYWORDS  
EST.  
SOURCE  
Plasmodium falciparum  
ORGANISM  
malaria parasite P. falciparum.  
REFERENCE  
1 (bases 1 to 432)  
AUTHORS  
Tang, K., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Pape, D., Marra, M., Hillier, L., Martin, J., Wylie, T., Dante, M., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Jentes, E., Ronko, I., Tsagarishvili, R., Belaygorod, L., Franklin, C., Carr, L., Gow, A., Maguire, L., Richey, J., Wadkins, J., Kennedy, S., Levinso, D., Waterston, R., Willson, R. and Sibley, D.

TITLE  
WashU Plasmodium EST Project  
JOURNAL  
Unpublished (2001)  
COMMENT  
Contact: L. David Sibley  
WashU Plasmodium EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

Library was constructed by Debopam Chakrabarti DNA sequencing by:  
Washington University Genome Sequencing Center for information on  
obtaining a clone please contact: L. David Sibley  
(sibley@borcim.wustl.edu), Washington University  
Seq primer: -400P from Gibco.  
Location/Qualifiers  
1. .432  
/organism="Plasmodium falciparum"  
/db\_xref="taxon:5833"  
/clone\_lib="Plasmodium falciparum 3D7 asexual cDNA"  
/lab\_host="DH10B (GeneHog, Invitrogen, Inc.)"  
/note="Vector: pBluescript SK plus; Site\_1: EcoRI; Site\_2:  
XhoI; Library was constructed by Debopam Chakrabarti.  
Total RNA samples were isolated from mixed stage  
saponin(0.1%) lysed P. falciparum 3D7 infected  
erythrocytes by the acidic guanidinium-phenol chloroform  
method. The poly A+ RNA was isolated by the polyA+-tract  
mRNA isolation system (Promega, WI) using streptavidin  
magnosphere particles. Directional cDNA libraries were  
constructed by oligo d(T) priming of poly(A)+ RNA (5mg)  
into EcoRI and XhoI sites of 1 ZapII vector using the Zap  
cDNA synthesis kit (Stratagene, CA). The average size of the  
cDNA inserts in the library was between 1.0 and 1.5kb.  
Clones were mass excised using the ExAssist helper phage  
(Stratagene), the phagemids were precipitated with PEG  
8000 and extracted with phenol/chloroform. Phagemid DNA  
was electroporated into DH10B cells."

FEATURES  
source

BASE COUNT 166 a 49 c 59 g 158 t

ORIGIN

Alignment Scores:  
Pred. No.: 4.29e-43 Length: 432  
Score: 484.00 Matches: 83  
Percent Similarity: 100.00% Conservations: 1  
Best Local Similarity: 98.81% Mismatches: 0  
Query Match: 21.27% Indels: 0  
DB: 14 Gaps: 0

US-10-057-531A-2 (1-431) x BQ633144 (1-432)

Qy 348 GlnCysProGluAsnSerGlyCysPheArgHisLeuAspGluArgGluGluCysLysCys 367  
|||||  
Db 3 CAATGTCAGAAAATTCGGATGTTTCAGACATTTAGTGAAGAGAGAGATGTAATGT 62  
|||||  
Qy 368 LeuLeuAsnTyrLysGlnGluGlyAspLysCysValGluAsnProAsnProThrCysAsn 387  
|||||  
Db 63 TTATTAATTTACAAACAAAGAGGTGATAAATGTTGTAATCCCAATCTACTGTGAAC 122  
|||||  
Qy 388 GluAsnAsnGlyGlyCysAspAlaAspAlaThrCysThrGluGluAspSerGlySerSer 407  
|||||

```

Db 123 GAAAAATAATGGTGATGTGATGCAGATGCCACATGTACCAAGAGATTACAGTACAGC 182
QY 408 ArqLysLysIleThrCysGluCysThrLysProAspSerTyrProLeuPheAspGlyIle 427
      |||||
Db 183 AGAAAGAAATACATGTGAATGTTCTAAACCTGATCTTATCCACTTTTCGATGTATT 242
QY 428 PheCysSerSer 431
      |||||
Db 243 TTCTGCAGTTC 254

RESULT 6
BI814631/c
LOCUS
DEFINITION
  BI814631
  PfEST0aa40c07.y1 Plasmodium falci-parum 3D7 asexual cDNA Plasmodium
  falci-parum cDNA 5' similar to TR:Q9UA18 Q9UA18 SURFACE PROTEIN-1 ;
  mRNA sequence.
ACCESSION
  BI814631
VERSION
  BI814631.1 GI:15907251
SOURCE
  malaria parasite P. falci-parum.
ORGANISM
  Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE
  1 (bases 1 to 501)
AUTHORS
  Tang,K., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Pape,D.,
  Marra,M., Hillier,L., Martin,J., Wyllie,T., Dente,M., Theising,B.,
  Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Jentes,E., Ronko,I.,
  Tsagarrelshvili,R., Belaygorod,L., Franklin,C., Carr,L., Grow,A.,
  Maguire,L., Richey,J., Wadkins,J., Kennedy,S., Levinso,D.,
  Waterston,R., Wilson,R. and Sibley,D.
  WashU Plasmodium EST Project
  Unpublished (2001)
  Contact: L. David Sibley
  WashU Plasmodium EST Project
  Washington University School of Medicine
  4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
  Tel: 314 286 1800
  Fax: 314 286 1810
  Email: est@watson.wustl.edu
  Library was constructed by Debopam Chakrabarti DNA sequencing by:
  Washington University Genome Sequencing Center For information on
  obtaining a clone please contact: L. David Sibley
  (sibley@borcim.wustl.edu), Washington University
  Possible reversed clone: similarity on wrong strand
  Seq primer: -40UP from Gibco
  High quality sequence stop: 261.

FEATURES
    source
    1..501
    /organism="Plasmodium falci-parum"
    /db_xref="taxon:5833"
    /clone_lib="Plasmodium falci-parum 3D7 asexual cDNA"
    /lab_host="DH10B (GensHog, Invitrogen, Inc.)"
    /note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2:
    XhoI; Library was constructed by Debopam Chakrabarti.
    Total RNA samples were isolated from mixed stage
    saponin(0.1%) lysed P. falci-parum 3D7 infected
    erythrocytes by the acidic guanidinium-phenol chloroform
    method. The poly A+ RNA was isolated by the polyAT-Tract
    mRNA isolation system (Promega, WI) using streptavidin
    Magnisphere particles. Directional cDNA libraries were
    constructed by oligo d(T) priming of poly(A)+ RNA (5mg)
    into EcoRI and XhoI sites of 1 ZapII vector using the Zap
    cDNA synthesis kit (Stragene, CA). The average size of the
    cDNA inserts in the library was between 1.0 and 1.5kb.
    Clones were mass excised using the ExAssist helper phage
    (Stratagene), the phagemids were precipitated with PEG
    8000 and extracted with phenol/chloroform. Phagemid DNA
    was electroporated into DH10B cells."
    181 a 62 c 51 g 207 t

BASE COUNT
  181 a 62 c 51 g 207 t

Alignment Scores:
  Pred. No.: 1.22e-34 Length: 501
  Score: 408.00 Matches: 71

```

```

Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 17.93% Indels: 0
DB: 13 Gaps: 0

US-10-057-531A-2 (1-431) x BI814631 (1-501)

QY 361 GluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLysCysValGlu 380
      |||||
Db 501 GAAAGAGAGAGATGTAATGTTTATTAAATTACAACACAGAGGTGATAATGTGTGAA 442
QY 381 AsnProAsnProThrCysAsnGluAsnGlyGlyCysAspAlaAspAlaThrCysThr 400
      |||||
Db 441 AATCCAAATCCTACTTGTACGAAATAATGGTGGATGTGATGCAGATGCCATGTACC 382
QY 401 GluGluAspSerCysSerArgLysLysIleThrCysGluCysThrLysProAspSer 420
      |||||
Db 381 GAAGAAGATTCAGTAGCAGACAGAAAGAAATCAGATGTGATGTACTAAACCTGATTCT 322
QY 421 TyrProLeuPheAspGlyIlePheCysSerSer 431
      |||||
Db 321 TATCCACTTTTCGATGTGATTTTCTGCAGTTC 289

RESULT 7
BQ451709
LOCUS
DEFINITION
  BQ451709
  PfEST0ab07h03.y1 Plasmodium falci-parum 3D7 asexual cDNA Plasmodium
  falci-parum cDNA 5' similar to TR:Q9UA18 Q9UA18 SURFACE PROTEIN-1 ;
  mRNA sequence.
ACCESSION
  BQ451709
VERSION
  BQ451709.1 GI:21254821
KEYWORDS
  EST.
SOURCE
  malaria parasite P. falci-parum.
ORGANISM
  Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE
  1 (bases 1 to 422)
AUTHORS
  Tang,K., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Pape,D.,
  Marra,M., Hillier,L., Martin,J., Wyllie,T., Dente,M., Theising,B.,
  Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Jentes,E., Ronko,I.,
  Tsagarrelshvili,R., Belaygorod,L., Franklin,C., Carr,L., Grow,A.,
  Maguire,L., Richey,J., Wadkins,J., Kennedy,S., Levinso,D.,
  Waterston,R., Wilson,R. and Sibley,D.
  WashU Plasmodium EST Project
  Unpublished (2001)
  Contact: L. David Sibley
  WashU Plasmodium EST Project
  Washington University School of Medicine
  4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
  Tel: 314 286 1800
  Fax: 314 286 1810
  Email: est@watson.wustl.edu
  Library was constructed by Debopam Chakrabarti DNA sequencing by:
  Washington University Genome Sequencing Center For information on
  obtaining a clone please contact: L. David Sibley
  (sibley@borcim.wustl.edu), Washington University
  Seq primer: -40UP from Gibco.
  Location/Qualifiers
    1..422
    /organism="Plasmodium falci-parum"
    /db_xref="taxon:5833"
    /clone_lib="Plasmodium falci-parum 3D7 asexual cDNA"
    /lab_host="DH10B (GensHog, Invitrogen, Inc.)"
    /note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2:
    XhoI; Library was constructed by Debopam Chakrabarti.
    Total RNA samples were isolated from mixed stage
    saponin(0.1%) lysed P. falci-parum 3D7 infected
    erythrocytes by the acidic guanidinium-phenol chloroform
    method. The poly A+ RNA was isolated by the polyAT-Tract
    mRNA isolation system (Promega, WI) using streptavidin
    Magnisphere particles. Directional cDNA libraries were
    constructed by oligo d(T) priming of poly(A)+ RNA (5mg)
    into EcoRI and XhoI sites of 1 ZapII vector using the Zap
    cDNA synthesis kit (Stragene, CA). The average size of the
    cDNA inserts in the library was between 1.0 and 1.5kb.
    Clones were mass excised using the ExAssist helper phage
    (Stratagene), the phagemids were precipitated with PEG
    8000 and extracted with phenol/chloroform. Phagemid DNA
    was electroporated into DH10B cells."

```

cDNA inserts in the library was between 1.0 and 1.5kb. Clones were mass excised using the EXAssist helper phage (Stratagene), the phagemids were precipitated with PEG 8000 and extracted with phenol/chloroform. Phagemid DNA was electroporated into DH10B cells."

BASE COUNT 162 a 44 c 57 g 159 t

ORIGIN

Alignment Scores:  
Pred. No.: 3,45e-34 Length: 422  
Score: 403.00 Matches: 70  
Percent Similarity: 100.00% Conservativeness: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 17.71% Indels: 0  
DB: 14 Gaps: 0

US-10-057-531A-2 (1-431) x BQ451709 (1-422)

Qy 362 ArgGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLysCysValGluAsn 381  
|||||  
Db 3 AGAGAAGAAATGTAATGTTTAAATACAAACAAGAGGTGATAATGTTGAAAT 62  
Qy 382 ProAsnProThrCysAsnGluAsnGlyCysAspAlaLaThrCysThrGlu 401  
|||||  
Db 63 CCAATCTACTTGAACGAAATTAATGGTGGATGATGCAGATGCCACATGTACCGAA 122  
Qy 402 GluAspSerGlySerArgLysLysIleThrCysGluCysThrLysProAspSerTyr 421  
|||||  
Db 123 GAACATTCAGTACGACGACAGAAATACATGTCATGTAATCTAAACCTGATTCATAT 182  
Qy 422 ProLeupheAspGlyIlePheCysSerSer 431  
|||||  
Db 183 CCACCTTTTCGATGTTATTTCTTCGAGTTC 212

RESULT 8  
BQ452445 522 bp mRNA linear EST 29-MAY-2002  
LOCUS  
DEFINITION  
PFEStoab09605.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium falciparum cDNA 5' similar to TR:Q9UA18 Q9UA18 SURFACE PROTEIN-1 ; mRNA sequence.

ACCESSION BQ452445  
VERSION BQ452445.1  
KEYWORDS GI:21255557

SOURCE EST.  
ORGANISM malaria parasite P. falciparum.

REFERENCE 1 (bases 1 to 522)  
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

AUTHORS  
Tang,K., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Pape,D., Marra,M., Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B., Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Jentes,E., Ronko,I., Tsagareishvili,R., Belaygorod,L., Franklin,C., Carr,L., Grow,A., Maguire,L., Richey,J., Wadkins,J., Kennedy,S., Levinso,D., Waterston,R., Willson,R. and Sibley,D.

TITLE WashU Plasmodium EST Project

JOURNAL Unpublished (2001)

COMMENT Contact: L. David Sibley

WashU Plasmodium EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.wustl.edu

Library was constructed by Debopam Chakrabarti DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: L. David Sibley

(sibley@borcim.wustl.edu), Washington University

Seq primer: -400P from Gibco

High quality sequence stop: 422.

Location/Qualifiers

FEATURES

source

1. .522 /organism="Plasmodium falciparum"

/db\_xref="taxon:5833"

/clone\_lib="Plasmodium falciparum 3D7 asexual cDNA"

/lab\_host="DH10B (GeneHog, Invitrogen, Inc.)"  
/note="Vector: pBluescript SK plus; Site\_1: EcoRI; Site\_2: XhoI; Library was constructed by Debopam Chakrabarti. Total RNA samples were isolated from mixed stage saponin(0.1%) lysed P. falciparum 3D7 infected erythrocytes by the acidic guanidinium-phenol chloroform method. The poly A+ RNA was isolated by the polyAT-Tract mRNA isolation system (Promega, WI) using streptavidin Magnosphere particles. Directional cDNA libraries were constructed by oligo d(T) priming of poly(A)+ RNA (5mg) into EcoRI and XhoI sites of 1 ZapII vector using the Zap cDNA synthesis kit (Stratagene, CA). The average size of the cDNA inserts in the library was between 1.0 and 1.5kb. Clones were mass excised using the EXAssist helper phage (Stratagene). the phagemids were precipitated with PEG 8000 and extracted with phenol/chloroform. Phagemid DNA was electroporated into DH10B cells."

BASE COUNT 204 a 63 c 68 g 187 t

ORIGIN

Alignment Scores:  
Pred. No.: 1,63e-33 Length: 522  
Score: 398.00 Matches: 69  
Percent Similarity: 100.00% Conservativeness: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 17.49% Indels: 0  
DB: 14 Gaps: 0

US-10-057-531A-2 (1-431) x BQ452445 (1-522)

Qy 363 GluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLysCysValGluAsnPro 382  
|||||  
Db 1 GAAGAATGTAATGTTTAAATACAAACAAGAGGTGATAATGTTGAAATCCA 60  
Qy 383 AsnProThrCysAsnGluAsnGlyCysAspAlaLaThrCysThrGluGlu 402  
|||||  
Db 61 AATCCTACTTGTAAACGAAATTAATGGTGGATGATGCAGATGCCACATGTACCGAAGAA 120  
Qy 403 AspSerGlySerArgLysLysIleThrCysGluCysThrLysProAspSerTyrPro 422  
|||||  
Db 121 GATTGAGTACGACGACGAAAGAAATCACATGTCATGTAATCAACCTGATTCATCCA 180  
Qy 423 LeupheAspGlyIlePheCysSerSer 431  
|||||

Db 181 CTTTTCGATGTTATTTCTTCGAGTTC 207

RESULT 9

BQ451036

LOCUS

DEFINITION

PFEStoab03a10.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium falciparum cDNA 5' similar to TR:Q9UA18 Q9UA18 SURFACE PROTEIN-1 ; mRNA sequence.

ACCESSION BQ451036

VERSION BQ451036.1

KEYWORDS GI:21254148

SOURCE EST.

ORGANISM malaria parasite P. falciparum.

REFERENCE 1 (bases 1 to 455)

Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

AUTHORS

Tang,K., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Pape,D., Marra,M., Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B., Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Jentes,E., Ronko,I., Tsagareishvili,R., Belaygorod,L., Franklin,C., Carr,L., Grow,A., Maguire,L., Richey,J., Wadkins,J., Kennedy,S., Levinso,D., Waterston,R., Willson,R. and Sibley,D.

TITLE WashU Plasmodium EST Project

JOURNAL Unpublished (2001)

COMMENT Contact: L. David Sibley

WashU Plasmodium EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.wustl.edu

Library was constructed by Debopam Chakrabarti DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: L. David Sibley

(sibley@borcim.wustl.edu), Washington University

Seq primer: -400P from Gibco

High quality sequence stop: 422.

Location/Qualifiers

FEATURES

source

1. .522 /organism="Plasmodium falciparum"

/db\_xref="taxon:5833"

/clone\_lib="Plasmodium falciparum 3D7 asexual cDNA"

Email: estwatson.wustl.edu  
 Library was constructed by Debopam Chakrabarti DNA sequencing by:  
 Washington University Genome Sequencing Center For information on  
 obtaining a clone please contact: L. David Sibley  
 (sibley@borcim.wustl.edu), Washington University  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 426.

#### FEATURES

source

1. 455

/organism="Plasmodium falciparum"  
 /db\_xref="taxon:5833"  
 /clone\_lib="Plasmodium falciparum 3D7 asexual cDNA"  
 /lab\_host="DH10B (GeneHog, Invitrogen, Inc.)"  
 /note="Vector: pBluescript SK plus; Site\_1: EcoRI; Site\_2:  
 XhoI; Library was constructed by Debopam Chakrabarti.  
 Total RNA samples were isolated from mixed stage  
 saponin(0.1%) lysed P. falciparum 3D7 infected  
 erythrocytes by the acidic guanidium-phenol chloroform  
 method. The poly A+ RNA was isolated by the polyAT-tract  
 mRNA isolation system (Promega, WI) using streptavidin  
 Magnosphere particles. Directional cDNA libraries were  
 constructed by oligo d(T) priming of poly(A)+ RNA (5mg)  
 into EcoRI and XhoI sites of 1 ZapII vector using the Zap  
 cDNA synthesis kit (Stragene, CA). The average size of the  
 cDNA inserts in the library was between 1.0 and 1.5kb.  
 Clones were mass excised using the ExAssist helper phage  
 (Stratagene), the phagemids were precipitated with PEG  
 8000 and extracted with phenol/chloroform. Phagemid DNA  
 was electroporated into DH10B cells."

179 a 46 c 56 g 174 t

#### BASE COUNT

ORIGIN

Alignment Scores:  
 Pred. No.: 4,82e-33 Length: 455  
 Score: 393.00 Matches: 68  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 17.27% Indels: 0  
 DB: 14 Gaps: 0

US-10-057-531A-2 (1-431) x BQ451036 (1-455)

QY 364 GLUCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLysCysValGluAsnProAsn 393  
 |||||  
 Db 3 GAATGTAATGTTTATTAAATTAACAACAGAGGTCATAAATGTTGTAATAATCCAAAT 62  
 QY 384 ProThrCysAsnGluAsnGlyGlyCysAspAlaAspAlaThrCysThrGluGluAsp 403  
 |||||  
 Db 63 CCTACTGTGTAACGAAATAATATGTTGGATGTGATGCAGATGCCACATGTACCGAAGAAGAT 122  
 QY 404 SerGlySerSerArgLysLysIleThrCysGluCysThrLysProAspSerTyrProLeu 423  
 |||||  
 Db 123 TCAGGTAGCAGCAGCAAGAAATACATCATGTGAATGTACTAAACCTGATTCTTATCCACTT 182  
 QY 424 PheAspGlyIlePheCysSerSer 431  
 |||||  
 Db 183 TTCGATGGTATTTTCTGCAGTTCC 206

#### RESULT 10

BI815239

LOCUS PfEST00a16b11.y1 Plasmodium falciparum 3D7 asexual cDNA.Plasmodium  
 DEFINITION falciparum cDNA 5' similar to TR:Q9UA18 Q9UA18 SURFACE PROTEIN-1 ;  
 mRNA sequence.

ACCESSION BI815239

VERSION BI815239.1 GI:15908372

KEYWORDS EST.

SOURCE malaria parasite P. falciparum.

ORGANISM Plasmodium falciparum

REFERENCE 1 (bases 1 to 303)

AUTHORS Tang,K., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Pape,D.,  
 Marra,M., Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B.,

#### TITLE JOURNAL COMMENT

Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Jentes,E., Ronko,I.,  
 Tsagarisvilli,R., Belaygorod,L., Franklin,C., Carr,L., Grow,A.,  
 Maguire,L., Richey,J., Wadhwa,J., Kennedy,S., Levinso,D.,  
 Waterston,R., Wilson,R. and Sibley,D.  
 WashU Plasmodium EST Project  
 Unpublished (2001)  
 Contact: L. David Sibley  
 WashU Plasmodium EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: estwatson.wustl.edu  
 Library was constructed by Debopam Chakrabarti DNA sequencing by:  
 Washington University Genome Sequencing Center For information on  
 obtaining a clone please contact: L. David Sibley  
 (sibley@borcim.wustl.edu), Washington University  
 putative full length read  
 vector to vector length is 433  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 257.

#### FEATURES

source

Location/Qualifiers

1. 303  
 /organism="Plasmodium falciparum"  
 /db\_xref="taxon:5833"  
 /clone\_lib="Plasmodium falciparum 3D7 asexual cDNA"  
 /lab\_host="DH10B (GeneHog, Invitrogen, Inc.)"  
 /note="Vector: pBluescript SK plus; Site\_1: EcoRI; Site\_2:  
 XhoI; Library was constructed by Debopam Chakrabarti.  
 Total RNA samples were isolated from mixed stage  
 saponin(0.1%) lysed P. falciparum 3D7 infected  
 erythrocytes by the acidic guanidium-phenol chloroform  
 method. The poly A+ RNA was isolated by the polyAT-tract  
 mRNA isolation system (Promega, WI) using streptavidin  
 Magnosphere particles. Directional cDNA libraries were  
 constructed by oligo d(T) priming of poly(A)+ RNA (5mg)  
 into EcoRI and XhoI sites of 1 ZapII vector using the Zap  
 cDNA synthesis kit (Stragene, CA). The average size of the  
 cDNA inserts in the library was between 1.0 and 1.5kb.  
 Clones were mass excised using the ExAssist helper phage  
 (Stratagene), the phagemids were precipitated with PEG  
 8000 and extracted with phenol/chloroform. Phagemid DNA  
 was electroporated into DH10B cells."

BASE COUNT 106 a 41 c 41 g 115 t

#### ORIGIN

Alignment Scores:  
 Pred. No.: 6,47e-25 Length: 303  
 Score: 317.00 Matches: 55  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 13.93% Indels: 0  
 DB: 13 Gaps: 0

US-10-057-531A-2 (1-431) x BI815239 (1-303)

QY 377 LysCysValGluAsnProAsnProThrCysAsnGluAsnGlyGlyCysAspAlaAsp 396  
 |||||  
 Db 3 AAATGTTGAAATCCAAATCTACTTGTACGAAATAATGTTGATGATGCAGAT 62  
 QY 397 AlaThrCysThrGluGluAspSerGlySerSerArgLysLysIleThrCysGluCysThr 416  
 |||||  
 Db 63 GCCACATGTACCGAAGAAGATTCAGGTAGCAGCAGCAAGAAATACATCATGTGAATGACT 122  
 QY 417 LysProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431  
 |||||

Db 123 AAACCTGATTCTTATCCACTTTTCGATGGTATTTTCTGCAGTTCC 167

#### RESULT 11

BI815239

LOCUS

DEFINITION EST566629 PyBS Plasmodium yoelii cDNA clone PYCUP75 5' end,  
 mRNA sequence.

BI815239

LOCUS

DEFINITION EST566629 PyBS Plasmodium yoelii cDNA clone PYCUP75 5' end,  
 mRNA sequence.

BI815239

LOCUS

DEFINITION EST566629 PyBS Plasmodium yoelii cDNA clone PYCUP75 5' end,  
 mRNA sequence.



Washington University Genome Sequencing Center For information on  
obtaining a clone please contact: L. David Sibley  
(sibley@borcim.wustl.edu), Washington University  
Putative full length read  
vector to vector length is 409  
Seq primer: -40UP from Gibco.

# FEATURES

source  
1. .408  
Location/Qualifiers  
/organism="Plasmodium falciparum"  
/db\_xref="taxon:5833"  
/clone\_lib="Plasmodium falciparum 3D7 asexual cDNA"  
/lab\_host="DH10B (GeneHog, Invitrogen, Inc.)"  
/note="Vector: pBluescript SK plus; Site\_1: EcoRI; Site\_2:  
XhoI; Library was constructed by Debopam Chakrabarti.  
Total RNA samples were isolated from mixed stage  
saponin(0.1%) lysed P. falciparum 3D7 infected  
erythrocytes by the acidic guanidinium-phenol chloroform  
method. The poly A+ RNA was isolated by the polyAT-Tract  
mRNA isolation system (Promega, WI) using streptavidin  
Magnesphere particles. Directional cDNA libraries were  
constructed by oligo d(T) priming of poly(A)+ RNA (5mg)  
into EcoRI and XhoI sites of 1 ZapII vector using the Zap  
cDNA synthesis kit (Stratagene, CA). The average size of the  
cDNA inserts in the library was between 1.0 and 1.5kb.  
Clones were mass excised using the ExAssist helper phage  
(Stratagene), the phagemids were precipitated with PEG  
8000 and extracted with phenol/chloroform. Phagemid DNA  
was electroporated into DH10B cells."

BASE COUNT 161 a 43 c 46 g 158 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 8.9e-21 Length: 408  
Score: 281.00 Matches: 49  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 12.35% Indels: 0  
DB: 13 Gaps: 0

US-10-057-531A-2 (1-431) x B1814931 (1-408)

QY 383 AsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAlaThrCysThrGluGlu 402  
|||||  
Db 3 AATCCTACTTGTAAACGAAATAATGTTGGATGTGATGCAGATGCCACATGTACCGAAGAA 62  
QY 403 AspSerGlySerArgLysLysIleThrCysGluCysThrLysProAspSerTyrPro 422  
|||||  
Db 63 GATTCAGGTAGCAGCAGAAAGAAATACATGTGATGTACTAAACCTGATCTTATCCA 122  
QY 423 LeuPheAspGlyIlePheCysSerSer 431  
|||||  
Db 123 CTTTTCGATGGTATTTTCTGCAGTTCC 149

RESULT 13  
BQ451572 418 bp mRNA linear EST 29-MAY-2002  
LOCUS pFESToab06b11.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium  
falciparum cDNA 5' similar to TR:Q9UA18 Q9UA18 SURFACE PROTEIN-1 ;  
mRNA sequence.

ACCESSION BQ451572  
VERSION BQ451572.1 GI:21254684  
KEYWORDS EST.

ORGANISM malaria parasite P. falciparum.

SOURCE Plasmodium falciparum

REFERENCE 1 (bases 1 to 418)  
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

AUTHORS Tang,K., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Pape,D.,  
Marra,M., Hillier,L., Martin,J., Wylie,T., Dente,M., Theising,B.,  
Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Jentes,E., Ronko,I.,  
Tsagarashvili,L., R., Belaygorod,L., Franklin,C., Carr,L., Graw,A.,  
Maguire,L., Richey,J., Wadkins,J., Kennedy,S., Levinso,D.,  
Waterston,R., Wilson,R. and Sibley,D.

TITLE WashU Plasmodium EST Project

# JOURNAL COMMENT

Unpublished (2001)  
Contact: L. David Sibley  
WashU Plasmodium EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Email: est@wustl.wustl.edu  
Library was constructed by Debopam Chakrabarti DNA sequencing by:  
Washington University Genome Sequencing Center For information on  
obtaining a clone please contact: L. David Sibley  
(sibley@borcim.wustl.edu), Washington University  
Seq primer: -40UP from Gibco.

# FEATURES

source

1. .418  
Location/Qualifiers  
/organism="Plasmodium falciparum"  
/db\_xref="taxon:5833"  
/clone\_lib="Plasmodium falciparum 3D7 asexual cDNA"  
/lab\_host="DH10B (GeneHog, Invitrogen, Inc.)"  
/note="Vector: pBluescript SK plus; Site\_1: EcoRI; Site\_2:  
XhoI; Library was constructed by Debopam Chakrabarti.  
Total RNA samples were isolated from mixed stage  
saponin(0.1%) lysed P. falciparum 3D7 infected  
erythrocytes by the acidic guanidinium-phenol chloroform  
method. The poly A+ RNA was isolated by the polyAT-Tract  
mRNA isolation system (Promega, WI) using streptavidin  
Magnesphere particles. Directional cDNA libraries were  
constructed by oligo d(T) priming of poly(A)+ RNA (5mg)  
into EcoRI and XhoI sites of 1 ZapII vector using the Zap  
cDNA synthesis kit (Stratagene, CA). The average size of the  
cDNA inserts in the library was between 1.0 and 1.5kb.  
Clones were mass excised using the ExAssist helper phage  
(Stratagene), the phagemids were precipitated with PEG  
8000 and extracted with phenol/chloroform. Phagemid DNA  
was electroporated into DH10B cells."

BASE COUNT 171 a 43 c 46 g 158 t  
ORIGIN

Alignment Scores:  
Pred. No.: 9.2e-21 Length: 418  
Score: 281.00 Matches: 49  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 12.35% Indels: 0  
DB: 14 Gaps: 0

US-10-057-531A-2 (1-431) x BQ451572 (1-418)

QY 383 AsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAlaThrCysThrGluGlu 402  
|||||  
Db 3 AATCCTACTTGTAAACGAAATAATGTTGGATGTGATGCAGATGCCACATGTACCGAAGAA 62

QY 403 AspSerGlySerArgLysLysIleThrCysGluCysThrLysProAspSerTyrPro 422  
|||||

Db 63 GATTCAGGTAGCAGCAGAAAGAAATACATGTGATGTACTAAACCTGATCTTATCCA 122  
QY 423 LeuPheAspGlyIlePheCysSerSer 431  
|||||  
Db 123 CTTTTCGATGGTATTTTCTGCAGTTCC 149

RESULT 14  
BQ451661

LOCUS pFESToab07c09.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium  
falciparum cDNA 5' similar to TR:Q9UA18 Q9UA18 SURFACE PROTEIN-1 ;  
mRNA sequence.

ACCESSION BQ451661

VERSION BQ451661.1 GI:21254773

KEYWORDS EST.

SOURCE malaria parasite P. falciparum.

ORGANISM Plasmodium falciparum

REFERENCE 1 (bases 1 to 422)  
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

**AUTHORS**  
 Tang, K., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Pape, D.,  
 Marra, M., Hillier, L., Martin, J., Wylie, T., Dante, M., Theising, B.,  
 Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Jentes, E., Ronko, I.,  
 Tsagaris, V., Belaygorod, L., Franklin, C., Carr, L., Grow, A.,  
 Maquire, L., Richey, J., Watkins, J., Kennedy, S., Levinso, D.,  
 Waterston, R., Wilson, R., and Sibley, D.  
**WashU Plasmodium EST Project**  
 Unpublished (2001)  
 Contact: L. David Sibley  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 Library was constructed by Debopam Chakrabarti DNA sequencing by:  
 Washington University Genome Sequencing Center For information on  
 obtaining a clone please contact: L. David Sibley  
 (sibley@borcim.wustl.edu), Washington University  
 Seq primer: -400P from Gibco.

**FEATURES**  
 source  
 1. 422  
 Location/Qualifiers  
 /organism="Plasmodium falciparum"  
 /db\_xref="taxon:5833"  
 /clone\_lib="Plasmodium falciparum 3D7 asexual cDNA"  
 /lab\_host="DH10B (GeneHog, Invitrogen, Inc.)"  
 /note="vector: pBluescript SK plus; Site\_1: EcoRI; Site\_2:  
 XhoI; Library was constructed by Debopam Chakrabarti.  
 Total RNA samples were isolated from mixed stage  
 saponin(0.1%) - lysed P. falciparum 3D7 infected  
 erythrocytes by the acidic guanidinium-phenol chloroform  
 method. The poly A+ RNA was isolated by the polyAT-tract  
 mRNA isolation system (Promega, WI) using streptavidin  
 Magnesphere particles. Directional cDNA libraries were  
 constructed by oligo d(T) priming of poly(A)+ RNA (5mg)  
 into EcoRI and XhoI sites of 1 ZapII vector using the Zap  
 cDNA synthesis kit (Stratagene, CA). The average size of the  
 cDNA inserts in the library was between 1.0 and 1.5kb.  
 Clones were mass excised using the EXassist helper phage  
 (Stratagene), the phagemids were precipitated with PEG  
 8000 and extracted with phenol/chloroform. Phagemid DNA  
 was electroporated into DH10B cells."

BASE COUNT 175 a 43 c 46 g 158 t

Alignment Scores:  
 Pred. No.: 9.32e-21 Length: 422  
 Score: 281.00 Matches: 49  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 12.35% Indels: 0  
 DB: 14 Gaps: 0

US-10-057-531a-2 (1-431) x BQ451661 (1-422)

Qy 383 AsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAaspAlaThrCysThrGluGlu 402  
 Db 3 AATCTACTTGTACGAAATAAATGCTGATGTCAGATGCCACATGTACCGAAGA 62

Qy 403 AspSerGlySerSerArgLysLysIleThrCysGluCysThrLysProAspSerTyrPro 422  
 Db 63 GATTGAGTACGACGAGAAAGAAATACATGTCGATGTAACCTGATCTTATCC 122

Qy 423 LeuPheAspGlyIlePheCysSerSer 431  
 Db 123 CTTTTCGATGGTATTTCTGCGAGTCC 149  
 RESULT 15  
 LOCUS BM167799 595 bp mRNA linear EST 04-DEC-2001  
 DEFINITION EST570322 PYBS Plasmodium yoelii cDNA clone PYCO81 5' end,  
 mRNA sequence.  
 ACCESSION BM167799

**VERSION**  
 BM167799.1 GI:17301031  
**KEYWORDS**  
 EST.  
**SOURCE**  
 Plasmodium yoelii yoelii.  
**ORGANISM**  
 Plasmodium yoelii yoelii  
**REFERENCE**  
 1 (bases 1 to 595)  
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
**AUTHORS**  
 Carlton, J.M., Daly, T.M., Long, C.A., Bergman, L.W., Vaidya, A.B.,  
 Fraser, C.M. and Carucci, D.J.  
**TITLE**  
 Plasmodium yoelii EST project at TIGR  
**JOURNAL**  
 Unpublished (2001)  
**COMMENT**  
 Contact: Jane Carlton  
 Parasite Genomics Group  
 The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-530-9319  
 Fax: 301-838-0208  
 Email: carlton@tigr.org  
 For clone info, please contact the Malaria Research and Reference  
 Reagent Resource Center, ATCC  
 http://www.malaria.mr4.org/mr4pages/index.html  
 Seq primer: ADF.

**FEATURES**  
 source  
 1. 595  
 Location/Qualifiers

/organism="Plasmodium yoelii yoelii"  
 /strain="17XL"  
 /db\_xref="taxon:73239"  
 /clone="PYCOQ81"  
 /clone\_lib="PYBS"  
 /dev\_stage="Asexual blood stages"  
 /lab\_host="E. coli XL-1 Blue"  
 /note="Vector: pAD-GAL4; At 20-25% parasitemia, blood was  
 collected from BALB/cByJ mice infected with Py17XL  
 parasites, and leukocytes removed by passage over  
 microcrystalline cellulose columns. Total RNA was  
 isolated using the guanidinium isothiocyanate method, and  
 mRNA isolated using oligo(dT)-cellulose chromatography.  
 First strand cDNA synthesis was completed using a 50-base  
 primer and reverse transcriptase in the presence of  
 5-methyl dCTP. After second strand synthesis, uneven  
 termini were treated with Pfu DNA polymerase and EcoRI  
 adaptors ligated to the blunt ends. The sample was cleaved  
 with XhoI and separated on a Sephacryl S-500 column.  
 Size-fractionated cDNA was precipitated and ligated to  
 HybrizAP arms directionally using EcoRI-XhoI cleaved arms.  
 After packaging, the phagemid vector (pAD-GAL4) was  
 excised from the HybrizAP vector and plasmid DNA  
 isolated."

BASE COUNT 254 a 64 c 111 g 166 t

Alignment Scores:  
 Pred. No.: 2.47e-20 Length: 595  
 Score: 279.00 Matches: 70  
 Percent Similarity: 54.85% Conservative: 43  
 Best Local Similarity: 33.98% Mismatches: 78  
 Query Match: 12.26% Indels: 16  
 DB: 13 Gaps: 7

US-10-057-531a-2 (1-431) x BM167799 (1-595)

Qy 231 LeuThrAsnIleGluThrLeuTyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIle 250  
 Db 5 CTTGAAGATCTTAAGGATTATATGAAACCGTAATAGTCAAGCAGAGAATATTCAGAA 64

Qy 251 AsnLeuLysAlaLysIleAsnAspCysAsnValGluLysAspGluAlaHisValLysIle 270  
 Db 65 GAATTACAAATAGACTTGATTAATAAATGAAAAAGCTGAATTTGAATATTATACA 124

Qy 271 ThrLysLeuSerAspLeuLysAlaIleAspLysIleAspLysLeuPheLysAsnProTyr 290  
 Db 125 AAAAATTTAGAAAAATACATAAAATTCACCAAAAACCTGACGAATTTGTAGAACATCCA 184

Qy 291 AspPheGluAlaIleLysLysLeuIleAsnAspThrLysLysAspMetLeuGlyLys 310



```

Db 185 GAA-----AATAATAAACACATAGCTTCAATAGCTTTAAACAAC 223
Qy 311 LeuLeuSerThrGlyLeuVal---GlnAsnPheProAsnThrIleIleSerLysLeuIle 329
Db 224 TTAATAAATCTGGTTAGTAGGAGGAGGTAATCAAGAAATATTAGCAAAATGCTT 283
Qy 330 GluGlyLysPheGlnAspMetLeuAsnIleSer---GlnHisGlnCysVal---LysLys 347
Db 284 AACATGCGATGGTATGATTTATTAGGTGTAGACCCCTAAACATGATGTGTTGATACAAGA 343
Qy 348 GlnCysProGluAsnSerGlyCysPheArgHisLeuAspGluArgGluGluCysLysCys 367
Db 344 GATATTCCTCAAAATGCTGGATGTTTATAGAGATGATATGCTACTGAAGAATGGAGATGT 403
Qy 368 LeuLeuAsnThrLysGln---GluGlyAspLysCysValGluAsnProAsnProThrCys 386
Db 404 TTATTAGCTTACAAAAAGGTGAAGGTATACATGTGTAGAAATAATATATCTTACTTGT 463
Qy 387 AsnGluAsnAsnGlyCysAspAlaAspAlaThrCysThrGluGluAspSerGlySer 406
Db 464 GATATCAACAATGCTGGATGTGCTCAACTGCTAGTTGTCAAAATGCGGAAAGTACGGAA 523
Qy 407 SerArgLysLysIle-----ThrCysGluCysThrLysProAspSerTyrProLeuPhe 424
Db 524 AATTCCAAAAAATATATATATATATATATATATATATATATATATATATATATATAT 576
Qy 425 AspGlyIlePheCysSer 430
Db 577 GAAGGTGTGATCTGTACT 594

```

```

RESULT 16
BM160302 527 bp mRNA linear EST 04-DEC-2001
LOCUS EST562825 PyBS Plasmodium yoelii yoelii cDNA clone PYCP50 5' end,
DEFINITION mRNA sequence.

```

```

ACCESSION BM160302.1 GI:17305983
VERSION Plasmodium yoelii yoelii.
KEYWORDS Plasmodium yoelii yoelii.
SOURCE Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
ORGANISM
REFERENCE 1 (bases 1 to 527)
AUTHORS Carlon,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Vaidya,A.B.,
Fraser,C.M. and Carucci,D.J.
TITLE Plasmodium yoelii EST project at TIGR
JOURNAL Unpublished (2001)
COMMENT Contact: Jane Carlon
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlon@tigr.org
For clone info, please contact the Malaria Research and Reference
Reagent Resource Center, ATCC
http://www.malaria.mr4.org/mr4pages/index.html
Seq primer: ADF.

```

```

FEATURES
source
1..527
Location/Qualifiers
1..527
/organism="Plasmodium yoelii yoelii"
/strain="17XL"
/db_xref="taxon:73239"
/clone="PYCP50"
/clone_lib="PyBS"
/dev_stage="Asexual blood stages"
/lab_host="E. coli XL-1 Blue"
/notes="Vector: PAD-GAL4; At 20-25% parasitemia, blood was
collected from BALB/cByJ mice infected with Py17XL
parasites, and leukocytes removed by passage over
microcrystalline cellulose columns. Total RNA was
isolated using the guanidinium isothiocyanate method, and
mRNA isolated using oligo(dT)-cellulose chromatography.

```

First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybriZAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (pAD-GAL4) was excised from the HybriZAP vector and plasmid DNA isolated."

BASE COUNT 199 a 55 c 92 g 181 t  
ORIGIN

#### Alignment Scores:

Pred. No.: 1-23e-19 Length: 527  
Score: 272.00 Matches: 55  
Percent Similarity: 61.72% Conservative: 24  
Best Local Similarity: 42.97% Mismatches: 45  
Query Match: 11.96% Indels: 4  
DB: 13 Gaps: 4

US-10-057-531a-2 (1-431) x BM160302 (1-527)

```

Qy 308 LeuGlyLysLeuLeuSerThrGlyLeuVal---GlnAsnPheProAsnThrIleIleSer 326
Db 29 TTAACAACACTTAATAAATCTGTTTGTAGGAGAGGTGAATCAAGAAATATTAGCA 88
Qy 327 LysLeuIleGluGlyLysPheGlnAspMetLeuAsnIleSer---GlnHisGlnCysVal 345
Db 89 AAAATGCTTAACATGGATGGTATGATTTATTAGGTGTAGACCCCTAAACATGTATGT 148
Qy 346 ---LysLysGlnCysProGluAsnSerGlyCysPheArgHisLeuAspGluArgGluGlu 364
Db 149 GATCACAGAGATATCTCTAAAATGCTGGATGTTTGTAGAGATGATATGCTACTGAAGA 208
Qy 365 CysLysCysLeuLeuAsnThrLysGln---GluGlyAspLysCysValGluAsnProAsn 383
Db 209 TGGAGATGTTTATTAGGTGTACAAAAAGGTGAAGGTAAATACATGTGTAGAAAAATAA 268
Qy 384 ProThrCysAsnGluAsnAsnGlyCysAspAlaAspAlaThrCysThrGluGluAsp 403
Db 269 CCTACTTGTGATATCAACAATGGTGGATGTGATCCAACTGCTAGTTGTCAAAAATGCCGA 328
Qy 404 SerGlySerArgLysLysLysIleThrCysGluCysThrLysProAspSerTyrProLeu 423
Db 329 AGTACGGAAAAATCCAAAAAATTTATATGTACATGTAAAGAACCAACCCCTAATGCATAT 388
Qy 424 PheAspGlyIlePheCysSerSer 431
Db 389 TATGAAGGTGTATTCTGTAGTTCT 412

```

#### RESULT 17

```

BM166681 784 bp mRNA linear EST 04-DEC-2001
LOCUS EST569204 PyBS Plasmodium yoelii yoelii cDNA clone PYCNV68 5' end,
DEFINITION mRNA sequence.

```

```

ACCESSION BM166681
VERSION BM166681.1 GI:17299913
KEYWORDS EST.
SOURCE Plasmodium yoelii yoelii.
ORGANISM Plasmodium yoelii yoelii.

```

```

REFERENCE 1 (bases 1 to 784)
AUTHORS Carlon,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Vaidya,A.B.,
Fraser,C.M. and Carucci,D.J.
TITLE Plasmodium yoelii EST project at TIGR
JOURNAL Unpublished (2001)
COMMENT Contact: Jane Carlon
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319

```



5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybrizAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (PAD-GAL4) was excised from the HybrizAP vector and plasmid DNA isolated."

BASE COUNT 159 a 41 c 72 g 167 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 1-32e-16 Length: 439  
 Score: 243.50 Matches: 43  
 Percent Similarity: 67.44% Conservative: 15  
 Best Local Similarity: 50.00% Mismatches: 27  
 Query Match: 10.70% Indels: 1  
 DB: 13 Gaps: 1

US-10-057-531A-2 (1-431) x BM160668 (1-439)

QY 347 LysGlnCysProGluAsnSerGlyCysPheArgHisLeuAspGluArgGluGluCysLys 366  
 Db 8 AGAGATATTCTCTAAATGCTGGATGTTTATAGAGATGATGATGCTACTGAAGAATGGAGA 67

QY 367 CysLeuLeuAsnTyrLysGln---GluGlyAspLysCysValGluAsnProAsnProThr 385  
 Db 68 TGTTTATTAGTTACAAAAGGTGAAGTAATACATGTGTGTAAGAAATAATAATATCTTACT 127

QY 386 CysAsnGluAsnGlyGlyCysAspAlaThrCysThrGluGluAspSerGly 405  
 Db 128 TGTGATATCAACAATGCTGGATGCTGATCAACTGCTAGTTGTCAAAATGCGGAAGTACG 187

QY 406 SerSerArgLysLysLeuThrCysGluCysThrLysProAspSerTyrProLeuPheAsp 425  
 Db 188 GAAATTTCCAAAAAATATATATGATGTAAGAACCAACCCCTAATGATATATGAA 247

QY 426 GlyIlePheCysSerSer 431  
 Db 248 GGTTATTTCTGTAGTTCT 265

RESULT 20  
 BI094652  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Laccaria bicolor.  
 Laccaria bicolor.  
 Agaricales; Tricholomataceae; Laccaria.  
 Podila, G.K., Brand, J.R. and Hynes, M.J.  
 Preinfection stage symbiosis-regulated cDNAs from L. bicolor x P. resinosa  
 Unpublished (2001)  
 Contact: Dr. G.K. Podila  
 Department of Biological Sciences  
 University of Alabama  
 Huntsville, AL 35899, USA  
 Tel: 256 824 6263  
 Fax: 256 824 6305  
 Email: podilag@uah.edu.  
 Location/Qualifiers  
 1..648  
 /organism="Laccaria bicolor"  
 /db\_xref="taxon:29883"  
 /clone\_lib="Preinfection stage symbiosis-regulated cDNAs from L. bicolor x P. resinosa"  
 /note="cDNA clones were selected using a combination of

FEATURES  
 source

271 ThrLysLeuSerAspLeuLysAlaIleAspLysIleAspLeuPheLysAsnProTyr 290  
 Db 125 AAAAAATTTACAAAATACATACAAATGACGAAATCTGACCAATTTGTAGAACATGCA 184

QY 291 AspPheGluAlaIleLysLysLeuIleAsnAspAspThrLysLysAspMetLeuGlyLys 310  
 Db 185 GAA-----AATAATAACACATAGCCTCAATAGCTTTAAACAAC 223

QY 311 LeuLeuSerThrGlyLeuVal---GlnAsnPheProAsnThrIleIleSerLysLeuIle 329  
 Db 224 TTAATAATAATCTGTTTGTAGTAGAGAGGTGAATCAAGAAATAATATAGCAAAATGCTT 283

QY 330 GluGlyLysPheGlnAspMetLeuAsnIleSer---GlnHisGlnCysVal---LysLys 347  
 Db 284 ACATGATGATGATGATGATTTATAGGTAGACCTTAACATGATGTTGTGATACAAGA 343

QY 348 GlnCysProGluAsnSerGlyCysPheArgHisLeuAspGluArgGluGluCysLysCys 367  
 Db 344 GATATTCTCTAAATGCTGGATGTTTATAGAGATGATAATGCTACTGAAGAATGAGATGT 403

QY 368 LeuLeuAsnTyrLysGln---GluGlyAspLysCysValGluAsnProAsnProThrCys 386  
 Db 404 TTATTAGTTACAAAAGGTGAAGTAATACATGTGTGTAAGAAATAATAATCTTACTTGT 463

QY 387 AsnGluAsnGlyGlyCysAspAlaThrCysThrGluGluAspSer 404  
 Db 464 GATATCAACAATGCTGGATGCTGATCCCACTGCTAGTTGTCAAAATGCGGAATCT 517

RESULT 19  
 BM160668  
 LOCUS  
 DEFINITION  
 EST563191 PYBS Plasmodium yoelii yoelii cDNA clone PYCJU25 5' end,  
 mRNA sequence.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Plasmodium yoelii yoelii.  
 Plasmodium yoelii yoelii.  
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 I (bases 1 to 439)  
 Carlton, J.M., Daly, T.M., Long, C.A., Bergman, L.W., Vaidya, A.B.,  
 Fraser, C.M. and Carucci, D.J.  
 Plasmodium yoelii EST project at TIGR  
 Unpublished (2001)  
 Contact: Jane Carlton  
 Parasite Genomics Group  
 The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-530-9319  
 Fax: 301-838-0208  
 Email: carlton@tigr.org  
 For clone info, please contact the Malaria Research and Reference  
 Reagent Resource Center, ATCC  
 http://www.malaria.mr4.org/mr4pages/index.html  
 Seq primer: ADP.

FEATURES  
 source

1..439  
 /organism="Plasmodium yoelii yoelii"  
 /strain="17XL"  
 /db\_xref="taxon:73239"  
 /clone="PYCJU25"  
 /clone\_lib="PYBS"  
 /dev\_stage="Asexual blood stages"  
 /lab\_host="E. coli XL-1 Blue"  
 /note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was  
 collected from BAUB/cbyJ mice infected with Py17XL  
 parasites, and leukocytes removed by passage over  
 microcrystalline cellulose columns. Total RNA was  
 isolated using the guanidinium isothiocyanate method, and  
 mRNA isolated using oligo(dT)-cellulose chromatography.  
 First strand cDNA synthesis was completed using a 50-base  
 primer and reverse transcriptase in the presence of

5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybrizAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (PAD-GAL4) was excised from the HybrizAP vector and plasmid DNA isolated."

BASE COUNT 159 a 41 c 72 g 167 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 1-32e-16 Length: 439  
 Score: 243.50 Matches: 43  
 Percent Similarity: 67.44% Conservative: 15  
 Best Local Similarity: 50.00% Mismatches: 27  
 Query Match: 10.70% Indels: 1  
 DB: 13 Gaps: 1

US-10-057-531A-2 (1-431) x BM160668 (1-439)

QY 347 LysGlnCysProGluAsnSerGlyCysPheArgHisLeuAspGluArgGluGluCysLys 366  
 Db 8 AGAGATATTCTCTAAATGCTGGATGTTTATAGAGATGATGATGCTACTGAAGAATGGAGA 67

QY 367 CysLeuLeuAsnTyrLysGln---GluGlyAspLysCysValGluAsnProAsnProThr 385  
 Db 68 TGTTTATTAGTTACAAAAGGTGAAGTAATACATGTGTGTAAGAAATAATAATATCTTACT 127

QY 386 CysAsnGluAsnGlyGlyCysAspAlaThrCysThrGluGluAspSerGly 405  
 Db 128 TGTGATATCAACAATGCTGGATGCTGATCAACTGCTAGTTGTCAAAATGCGGAAGTACG 187

QY 406 SerSerArgLysLysLeuThrCysGluCysThrLysProAspSerTyrProLeuPheAsp 425  
 Db 188 GAAATTTCCAAAAAATATATATGATGTAAGAACCAACCCCTAATGATATATGAA 247

QY 426 GlyIlePheCysSerSer 431  
 Db 248 GGTTATTTCTGTAGTTCT 265

RESULT 20  
 BI094652  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Laccaria bicolor.  
 Laccaria bicolor.  
 Agaricales; Tricholomataceae; Laccaria.  
 Podila, G.K., Brand, J.R. and Hynes, M.J.  
 Preinfection stage symbiosis-regulated cDNAs from L. bicolor x P. resinosa  
 Unpublished (2001)  
 Contact: Dr. G.K. Podila  
 Department of Biological Sciences  
 University of Alabama  
 Huntsville, AL 35899, USA  
 Tel: 256 824 6263  
 Fax: 256 824 6305  
 Email: podilag@uah.edu.  
 Location/Qualifiers  
 1..648  
 /organism="Laccaria bicolor"  
 /db\_xref="taxon:29883"  
 /clone\_lib="Preinfection stage symbiosis-regulated cDNAs from L. bicolor x P. resinosa"  
 /note="cDNA clones were selected using a combination of

FEATURES  
 source

Suppressive subtraction hybridization derived probes and a cDNA library of L. bicolor, prepared from pooled RNA of interaction time points ranging from 6-72 h. Differential expression of these cDNAs was confirmed through screening a membrane array of the cDNAs with exponential probes prepared from control RNA obtained from free living L. bicolor and pooled interaction RNA."

BASE COUNT 175 a 149 c 182 g 142 t  
ORIGIN

Alignment Scores: 7.97e-16 Length: 648  
Pred. No.: 238.50 Matches: 49  
Score: 80.33% Conservative: 0  
Percent Similarity: 80.33% Mismatches: 1  
Best Local Similarity: 10.48% Indels: 12  
Query Match: 10.48% Gaps: 1  
DB: 13

US-10-057-531A-2 (1-431) x BI094652 (1-648)

Qy 1 MethHisHisHisHisHisSerSerGlyLeuValProArgGlySerGlyMetLysGlu 20  
|||||  
Db 338 ATGCACCATCATCATCTCTCTGGTCTGGTCCACCGGTTCTGGTATGAAGAA 397  
|||||

Qy 21 ThrAlaAlaLysPheGluArgGlnHisMetAspSerPro----- 34  
|||||  
Db 398 ACCGCTGCTGCTAAATTCGAACGCCAGCACATGGACAGCCCATCTGCATTTAGGTGAC 457  
|||||

Qy 35 -----AspLeuGlyThrAspAspAspLysAlaMetAlaAspIleGly 49  
|||||  
Db 458 ACTATAGTAATACCAAGATCTGGGTACCGACGACGACCAAGGCCATCGGC-GATATCGGA 516  
|||||

Qy 50 Ser 50  
|||  
Db 517 TCC 519

RESULT 21  
BI094593 661 bp mRNA linear EST 20-JUN-2001  
LOCUS  
DEFINITION LBSH00024 Preinfection stage symbiosis-regulated cDNAs from L. bicolor x P. resinosa Laccaria bicolor cDNA, mRNA sequence.

ACCESSION BI094593  
VERSION BI094593.1 GI:14516071  
KEYWORDS EST.  
SOURCE Laccaria bicolor.  
ORGANISM Laccaria bicolor.

REFERENCE 1 (bases 1 to 661)  
AUTHORS Podilla,G.K., Brand,J.R. and Hymes,M.J.  
TITLE Preinfection stage symbiosis-regulated cDNAs from L. bicolor x P. resinosa  
JOURNAL Unpublished (2001)  
COMMENT Contact: Dr. G.K. Podilla  
Department of Biological Sciences  
University of Alabama  
Huntsville, AL 35899, USA  
Tel: 256 824 6263  
Fax: 256 824 6305  
Email: podillag@mail.uah.edu.

FEATURES  
source  
1. .661  
/organism="Laccaria bicolor"  
/db\_xref="taxon:29883"  
/clone\_lib="Preinfection stage symbiosis-regulated cDNAs from L. bicolor x P. resinosa"  
/note="cDNA clones were selected using a combination of suppressive subtraction hybridization derived probes and a cDNA library of L. bicolor, prepared from pooled RNA of interaction time points ranging from 6-72 h. Differential expression of these cDNAs was confirmed through screening a membrane array of the cDNAs with exponential probes prepared from control RNA obtained from free living L.

BASE COUNT 183 a 145 c 180 g 153 t  
ORIGIN

Alignment Scores: 8.19e-16 Length: 661  
Pred. No.: 238.50 Matches: 49  
Score: 80.33% Conservative: 0  
Percent Similarity: 80.33% Mismatches: 1  
Best Local Similarity: 10.48% Indels: 12  
Query Match: 10.48% Gaps: 1  
DB: 13

US-10-057-531A-2 (1-431) x BI094593 (1-661)

Qy 1 MethHisHisHisHisHisSerSerGlyLeuValProArgGlySerGlyMetLysGlu 20  
|||||  
Db 412 ATGCACCATCATCATCTCTCTGGTCTGGTCCACCGGTTCTGGTATGAAGAA 471  
|||||

Qy 21 ThrAlaAlaLysPheGluArgGlnHisMetAspSerPro----- 34  
|||||  
Db 472 ACCGCTGCTGCTAAATTCGAACGCCAGCACATGGACAGCCCATCTGCATTTAGGTGAC 531  
|||||

Qy 35 -----AspLeuGlyThrAspAspAspLysAlaMetAlaAspIleGly 49  
|||||  
Db 532 ACTATAGTAATACCAAGATTTGGGTACCGACGACGACCAAGGCCATCGGC-GATATCGGA 590  
|||||

Qy 50 Ser 50  
|||  
Db 591 TCC 593

RESULT 22  
BI094577 686 bp mRNA linear EST 20-JUN-2001  
LOCUS

DEFINITION LBSH00008 Preinfection stage symbiosis-regulated cDNAs from L. bicolor x P. resinosa Laccaria bicolor cDNA, mRNA sequence.

ACCESSION BI094577  
VERSION BI094577.1 GI:14516055  
KEYWORDS EST.  
SOURCE Laccaria bicolor.  
ORGANISM Laccaria bicolor.

REFERENCE 1 (bases 1 to 686)  
AUTHORS Podilla,G.K., Brand,J.R. and Hymes,M.J.  
TITLE Preinfection stage symbiosis-regulated cDNAs from L. bicolor x P. resinosa  
JOURNAL Unpublished (2001)  
COMMENT Contact: Dr. G.K. Podilla  
Department of Biological Sciences  
University of Alabama  
Huntsville, AL 35899, USA  
Tel: 256 824 6263  
Fax: 256 824 6305  
Email: podillag@mail.uah.edu.

FEATURES  
source  
1. .686  
/organism="Laccaria bicolor"  
/db\_xref="taxon:29883"  
/clone\_lib="Preinfection stage symbiosis-regulated cDNAs from L. bicolor x P. resinosa"  
/note="cDNA clones were selected using a combination of suppressive subtraction hybridization derived probes and a cDNA library of L. bicolor, prepared from pooled RNA of interaction time points ranging from 6-72 h. Differential expression of these cDNAs was confirmed through screening a membrane array of the cDNAs with exponential probes prepared from control RNA obtained from free living L.

BASE COUNT 194 a 154 c 179 g 157 t  
ORIGIN

Alignment Scores: 8.61e-16 Length: 686  
Pred. No.: 238.50 Matches: 49  
Score: 80.33% Conservative: 0  
Percent Similarity: 80.33% Mismatches: 1  
Best Local Similarity: 10.48% Indels: 12  
Query Match: 10.48% Gaps: 1  
DB: 13

Score: 238.50 Matches: 49  
 Percent Similarity: 80.33% Conservative: 0  
 Best Local Similarity: 80.33% Mismatches: 1  
 Query Match: 10.48% Indels: 12  
 DB: 13 Gaps: 1

US-10-057-531A-2 (1-431) x BI094577 (1-686)

QY 1 MethHisHisHisHisHisSerGlyLeuValProArgGlySerGlyMetLysGlu 20

Db 406 ATGCACCATCATCATCATCTCTCTGGTCTGCTGCCAGCGGTCTCTGGTATGAAGAA 465

QY 21 ThrAlaAlaLysPheGluArgGlnHisMetAspSerPro----- 34

Db 466 ACCGCTGCTCTAAATTCGAACCCAGCACATGGACAGCCAGCATCTGCATTTAGGTGAC 525

QY 35 -----AspLeuGlyThrAspAspAspLysAlaMetAlaAspIleGly 49

Db 526 ACTATAGAATACCAAGATCTGGTACCGAGCAGACAAAGCCATGGC-GATATCGGA 584

QY 50 Ser 50

Db 585 TCC 587

RESULT 23

BI094637

LOCUS

DEFINITION

ACCESION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

Location/Qualifiers

1. .415

/organism="Laccaria bicolor"

/db\_xref="taxon:29883"

/clone\_lib="Preinfection stage symbiosis-regulated cDNAs

from L. bicolor x P. resinosa"

/note="cDNA clones were selected using a combination of

suppressive subtraction hybridization derived probes and

a cDNA library of L. bicolor, prepared from pooled RNA of

interaction time points ranging from 6-72 h. Differential

expression of these cDNAs was confirmed through screening

a membrane array of the cDNAs with exponential probes

prepared from control RNA obtained from free living L.

bicolor and pooled interaction RNA."

BASE COUNT 125 a 95 c 108 g 86 t 1 others

ORIGIN

Alignment Scores:

Pred. No.: 1.2e-15

Score: 234.50

Percent Similarity: 78.69%

Best Local Similarity: 78.69%

Query Match: 10.31%

DB: 13

Matches: 48

Conservative: 0

Mismatches: 2

Indels: 12

Gaps: 1

US-10-057-531A-2 (1-431) x BI094637 (1-415)

QY 1 MethHisHisHisHisHisSerGlyLeuValProArgGlySerGlyMetLysGlu 20

Db 158 ATGCACCATCATCATCATCTCTCTGGTCTGCTGCCAGCGGTCTCTGGTATGAAGAA 217

QY 21 ThrAlaAlaLysPheGluArgGlnHisMetAspSerPro----- 34

Db 218 ACCGCTGCTCTAAATTCGAACCCAGCACATGGACAGCCAGCATCTGCATTTAGGTGAC 277

QY 35 -----AspLeuGlyThrAspAspAspLysAlaMetAlaAspIleGly 49

Db 278 ACTATAGAATACCAAGATCTGGTACCGAGCAGACAAAGCCATGGC-GATATCGGA 336

QY 50 Ser 50

Db 337 TCC 339

RESULT 24

BF298872

LOCUS

DEFINITION

ACCESION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

Location/Qualifiers

1. .660

/organism="Plasmodium berghei"

/strain="ANKA clone 15cyl (clone of the ANKA 8417 HP clone

)"

/db\_xref="taxon:5821"

/clone\_lib="PB cDNA #20, Charles Yowell and Jane Carlton"

/dev\_stage="asynchronous blood stage"

/lab\_host="Swiss white mice"

/note="Vector: pBluescript II vector DNA, excised from

Lamda ZAP II.; Site.1: EcoRI; Site.2: XhoI; Total RNA was

extracted from asynchronous blood stage forms of the

Swiss white mice. Contaminating host white cells had

previously been removed using a novel biomagnetic bead

protocol (J. Carlton et al., manuscript in preparation).

PolyA+ RNA was extracted and reverse transcribed using an

oligo dT-XhoI primer. Second strand cDNA was prepared

using RNase H and DNA polymerase I. EcoRI adaptors were

ligated to the cDNA, and it was digested with XhoI.

Fragments were size selected, and those between 1-5 kb

ligated into EcoRI /XhoI digested vector."

BASE COUNT 307 a 71 c 99 g 181 t 2 others

ORIGIN

Alignment Scores:

Pred. No.: 2.25e-15

Score: 234.50

Percent Similarity: 51.65%

Best Local Similarity: 24.79%

Query Match: 10.31%

DB: 12

Length: 660

Matches: 60

Conservative: 65

Mismatches: 85

Indels: 32

Gaps: 5

US-10-057-531a-2 (1-431) x BF298872 (1-660)

Qy 111 ArgLeuLysLysArgLysTyrPheLeuAspValLeuGluSerAspLeuMetClnPhelLys 130  
 Db 1 AAATACAAAGAAAGAAATGATTCTCTGAAGTATTAAATCATGAATGATTCATCAAA 60  
 Qy 131 HisLysSerAsnGluTyrLysLeuGluAspSerPheLysLeuLeuAsnSerGluGln 150  
 Db 61 GATTTAAGTACCACAATATGTTATTAGAAACCATACCAATATTAGATAATGATAAA 120  
 Qy 151 LysAsnThrLeuLysSerTyrLysTyrLysLysGluSerValGluAsnAspLys 170  
 Db 121 AAAGACAACAATAAGTAAATTTAAATATGCTGCTAAAGGTGTAATCAAGATATAGAA 180  
 Qy 171 PheAlaGlnGluGlyLysSerTyrTyrGluLysValLeuAlaLysTyrLysAspLeu 190  
 Db 181 ACAACTGCTGACGGAATTAATCTTTAACAAAGATGATTGAATATACAAATCAATTA 240  
 Qy 191 GluSerLysLysValLysLysGluLysGluLysPheProSerSerProThr 210  
 Db 241 GCTGAGTAAAGAAACAATGTATGCTCATAGAA----- 273  
 Qy 211 ThrProSerProAlaLysThrAspGluLysLysGluSerLysPheLeuProPhe 230  
 Db 274 -----GCTGCTACTACCGATAAGATGAAAGAAAGAAATATGTTCCAATC 318  
 Qy 231 LeuThrAsnIleGluThrLysAsnAsnLeuValAsnLysIleAspAspTyrLys 250  
 Db 319 TTGAGATCTTAAGAGTATTATGAACCATATGATGATCAATCAGAGAATATATAGAA 378  
 Qy 251 AsnLeuLysAlaLysIleAsnAspCysAsnValGluLysAspGluAlaHisValLys 270  
 Db 379 GTATTACAAATAGACTTGATAGTTATATAAATGAAAGACGTAATNTGAAATTTAACG 438  
 Qy 271 ThrLysLeuSerAspLeuLysAlaLysAspLysIleAspLeuPheLysAsnProTyr 290  
 Db 439 AAAATTGTAGAACATACATAAAATGACGAAACCTTGAAATTTCT----- 486  
 Qy 291 AspPheGluAlaIleLysLysLysLeuAsnAspThrLysLysAspMet----- 307  
 Db 487 -----GTAGAAATGCAGAAATATAAATACATACATAGCCCTCAATG 525  
 Qy 308 ---LeuGlyLysLeuLysSerThrGlyLeuVal---GlnAsnPheProAsnThrLys 325  
 Db 526 GCTTTAAATAACTTAATAAATCTGTTGTGAGGAGGCGAATCGAAATAATATTA 585  
 Qy 326 SerLysLeuIleGluLysPheGlnAspMetLeuAsnIleSer---GlnHisGlnCys 344  
 Db 586 GCAAAATGCTTAATATGATAGTATGATTTATTAAATAGTAGACCTTAAGCATGTTG 645  
 Qy 344 sval 345  
 Db 646 TTTA 649

RESULT 25  
 BI094651  
 LOCUS  
 DEFINITION L85SH00094 Preinfection stage symbiosis-regulated cDNAs from L.  
 bicolor x P. resinosa Laccaria bicolor cDNA, mRNA sequence.  
 ACCESSION BI094651  
 VERSION BI094651.1 GI:14516128  
 KEYWORDS EST.  
 SOURCE Laccaria bicolor.  
 ORGANISM Laccaria bicolor  
 Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;  
 Agaricales; Tricholomataceae; Laccaria.  
 REFERENCE 1 (bases 1 to 405)  
 AUTHORS Podila,G.K., Brand,J.R. and Hynes,M.J.  
 TITLE Preinfection stage symbiosis-regulated cDNAs from L. bicolor x P.  
 resinosa  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Dr. G.K. Podila

Department of Biological Sciences  
 University of Alabama  
 Huntsville, AL 35899, USA  
 Tel: 256 824 6263  
 Fax: 256 824 6305  
 Email: podilag@mail.uah.edu.

FEATURES  
 source  
 1. 405  
 /organism="Laccaria bicolor"  
 /db\_xref="taxon:29883"  
 /clone\_lib="Preinfection stage symbiosis-regulated cDNAs  
 from L. bicolor x P. resinosa"  
 /note="cDNA clones were selected using a combination of  
 suppressive subtraction hybridization derived probes and  
 a cDNA library of L. bicolor, prepared from pooled RNA of  
 interaction time points ranging from 6-72 h. Differential  
 expression of these cDNAs was confirmed through screening  
 a membrane array of the cDNAs with exponential probes  
 prepared from control RNA obtained from free living L.  
 bicolor and pooled interaction RNA."

BASE COUNT 116 a 92 c 112 g 85 t  
 ORIGIN

Alignment Scores:  
 Pred. NO.: 1.5e-15 Length: 405  
 Score: 233.50 Matches: 56  
 Percent Similarity: 53.51% Conservative: 5  
 Best Local Similarity: 49.12% Mismatches: 12  
 Query Match: 10.26% Indels: 42  
 DB: 13 Gaps: 3

US-10-057-531a-2 (1-431) x BI094651 (1-405)

Qy 1 MethHisHisHisHisHisSerSerGlyLeuValProArgGlySerGlyMetLysGlu 20  
 Db 126 ATGCACCATCATCATCTCTCTGTCGTCGCCACCGCGTGTGGTATGAAGAA 185  
 Qy 21 ThrAlaLalaLysPheGluArgGlnHisMetAspSerPro----- 34  
 Db 186 ACCGCTGCTGCTAAATTCGAACGCCAGCACATGGACGCCAGATCTGCATTAGTGAC 245  
 Qy 35 -----AspLeuGlyThrAspAspAspLysAlaMetAlaAspLysGly 49  
 Db 246 CCTATAGAATACCAAGATCTGGTACCGACGACGACGACGATGGC-GATATCGGA 304  
 Qy 50 SerIleGluGlyArgGlyThrMetAlaIleSerValThrMetAspAsnIleLeuSerGly 69  
 Db 305 TCC----- 307  
 Qy 70 PheGluAsnGluTyrAspValIleTyrLeuLys-----Pro 81  
 Db 308 -----GAATTCGCGCGCGCTCGATACGGCTGCGAGAAGACGACAGAGGGTCCC 358  
 Qy 82 LeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLys 95  
 Db 359 CTTCGGGTGACATACGAAAAA----- 400

RESULT 26  
 BI094651  
 LOCUS  
 DEFINITION EST562555 PyBS Plasmodium yoelii yoelii cDNA clone PYCJM10 5' end,  
 mRNA sequence.  
 ACCESSION BI094651  
 VERSION BI094651.1 GI:17305713  
 KEYWORDS EST.  
 SOURCE Plasmodium yoelii yoelii.  
 ORGANISM Plasmodium yoelii yoelii  
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 REFERENCE 1 (bases 1 to 725)  
 AUTHORS Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Vaidya,A.B.,  
 Fraser,C.M. and Carucci,D.J.  
 TITLE Plasmodium yoelii EST project at TIGR  
 JOURNAL Unpublished (2001)







Interaction time points ranging from 6-72 h. Differential expression of these cDNAs was confirmed through screening a membrane array of the cDNAs with exponential probes prepared from control RNA obtained from free living L. bicolor and pooled interaction RNA."

BASE COUNT 168 a 125 c 152 g 126 t  
ORIGIN

Alignment Scores: 1.81e-14 Length: 571  
Pred. No.: 225.50 Matches: 49  
Score: 223.00  
Percent Similarity: 79.03% Conservative: 0  
Best Local Similarity: 79.03% Mismatches: 1  
Query Match: 9.91% Indels: 13  
DB: 13 Gaps: 1

US-10-057-531A-2 (1-431) x BI094610 (1-571)

QY 1 MetHisHisHisHisSerGlyLeuValProArg-GlySerGlyMetLysG1 20  
|||||  
Db 202 ATGCACCATCATCATCTCTCTGCTGCTGCCAGCGGTTCTGGTATGAAGA 261  
QY 20 uThrAlaAlaLysPheGluArgGlnHisMetAspSerPro----- 34  
|||||  
Db 262 AACCGCTGCTGCTAAATTCGACGCCAGCACATGGACGCCGATCTGTCATTTAGGTGA 321  
QY 35 -----AspLeuGlyThrAspAspAspLysAlaMetAlaAspLysG1 49  
|||||  
Db 322 CACTATAGATACCATGATCTGGTACCGACGACGACGACGACGATGCG-CATATCGG 380  
QY 49 ySer 50  
|||||  
Db 381 ATCC 384

#### RESULT 30

BM163150 619 bp mRNA linear EST 04-DEC-2001  
LOCUS EST565673 PyBS Plasmodium yoelii cDNA clone PYCLB55 5' end,  
DEFINITION mRNA sequence.

ACCESSION BM163150

VERSION BM163150.1 GI:17308831

KEYWORDS EST.

SOURCE Plasmodium yoelii yoelii.

ORGANISM Plasmodium yoelii yoelii

REFERENCE 1 (bases 1 to 619); Apicomplexa; Haemosporida; Plasmodium.

AUTHORS Carleton, J.M., Daly, T.M., Long, C.A., Bergman, L.W., Vaidya, A.B.,

Fraser, C.M. and Garucci, D.J.

Plasmodium yoelii EST project at TIGR

Unpublished (2001)

Contact: Jane Carleton

Parasite Genomics Group

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-530-9319

Fax: 301-838-0208

Email: carleton@tigr.org

For clone info, please contact the Malaria Research and Reference

Reagent Resource Center, ATCC

http://www.malaria.mr4.org/mr4pages/index.html

Seq primer: ADP.

Location/Qualifiers

1. .619

/organism="plasmodium yoelii yoelii"

/strain="17XL"

/db\_xref="taxon:73239"

/clone="PYCLB55"

/clone\_lib="PyBS"

/lab\_host="E. coli XL-1 Blue"

/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was

collected from BAUB/cbyJ mice infected with Py17XL

parasites, and leukocytes removed by passage over

microcrystalline cellulose columns. Total RNA was isolated using the guanidinium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column.

Size-fractionated cDNA was precipitated and ligated to HybrizAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (PAD-GAL4) was excised from the HybrizAP vector and plasmid DNA isolated."

BASE COUNT 257 a 99 c 115 g 148 t  
ORIGIN

#### Alignment Scores:

Pred. No.: 3.81e-14 Length: 619  
Score: 223.00 Matches: 57  
Percent Similarity: 51.38% Conservative: 72  
Best Local Similarity: 31.49% Mismatches: 76  
Query Match: 9.80% Indels: 16  
DB: 13 Gaps: 4

US-10-057-531A-2 (1-431) x BM163150 (1-619)

QY 10 GlyLeuValProArgGlySerGlyMetLysGluThrAlaAlaLysPheGluArgGln 29  
|||||  
Db 117 GGTGCGTACCA---GGATCAGGACCGGATACACGGGTAGCTGGA----- 158  
QY 30 HisMetAspSerProAspLeuGlyThrAspAspAspLysAlaMetAlaAspLysGly 49  
|||||  
Db 159 -----AGCAGTCTTGATGATATGAAGACGATCATATATCAATGCAAGT---GGT 209  
QY 50 SerIleGluGlyArgGlyThrMetAlaIleSerValThrMetAspAsnIleLeuSerGly 69  
|||||  
Db 210 CAATCCGAA-----GATGCACCCAGAAAAGATATTGTTCCGAA 248  
QY 70 PheGluAsnGluTyrAspValIleTyrLeuLysProLeuAlaGlyValTyrArgSerLeu 89  
|||||  
Db 249 TTTACAAATGAAGTTGTTGATGATATACAAAGGTTGGGTAGTACATATAAATCAATTA 308  
QY 90 LysLysGlnIleGluLysAsnIlePheThrPheAsnLeuAsnLeuAsnAspIleLeuAsn 109  
|||||  
Db 309 AAGAAACACATGTTAAGAGAATTTTCAACAATTAAGACACATGACAAATGGATTAAAT 368  
QY 110 SerArgLeuLysLysArgLysTyrPheLeuAspValLeuGluSerAspLeuMetGlnPhe 129  
|||||  
Db 369 AATAAATACAAAAAGAAATGATTTCTTGAAGTATTAAAGCCCATGAATTAGATTATTC 428  
QY 130 LysHisIleSerSerAsnGluTyrIleIleGluAspSerPheLysLeuLeuAsnSerGlu 149  
|||||  
Db 429 AAAGATTTAAGTACCAACAATATGTTATTAGAATCCATATCAATATTAGATAATGAT 488  
QY 150 GlnLysAsnThrLeuLeuLysSerTyrLysTyrIleLysGluSerValGluAsnAspIle 169  
|||||  
Db 489 AAAAAAGACAAACAAATAGTAAACTTTAAATATGCTACTAAAGGTATAAATGGAGATATA 548  
QY 170 LysPheAlaGlnGluGlyIleSerTyrTyrGluLysValLeuAlaLysTyrLysAspAsp 189  
|||||  
Db 549 GAAACAACTACTGACGGAATTAATTTCTTTAACAAAATGTTTGAATTTATACACACCTCA 608  
QY 190 Leu 190  
|||||  
Db 609 TTA 611

#### RESULT 31

BM164097

LOCUS

DEFINITION

ACCESION

BM164097 782 bp mRNA linear EST 04-DEC-2001  
EST566620 PyBS Plasmodium yoelii cDNA clone PYCLP63 5' end,  
mRNA sequence.  
BM164097

[illegible]

After packaging, the phagemid vector (pAD-GAL4) was excised from the HybridZAP vector and plasmid DNA isolated."

```

BASE COUNT      128 a   40 c   67 g   94 t
ORIGIN

Alignment Scores:
Pred. No.:      1.96e-11      Length:      329
Score:          195.00      Matches:      43
Percent Similarity: 62.11%      Conservative: 16
Best Local Similarity: 45.26%      Mismatches:  32
Query Match:      8.57%      Indels:       4
DB:              13      Gaps:         4

US-10-057-531A-2 (1-431) x BM167649 (1-329)

QY 308 LeuGlyLysLeuLeuSerThrGlyLeuVal---GlnAsnPheProAsnThrIleLeuSer 326
Db 43 TTAACAACCTTAATAATCTGGTTTAGTAGGAGAGAGTGAATCAAGAAATAATTAGCA 102
QY 327 LysLeuIleGluGlyLysPheGlnAspMetLeuAsnIleSer---GlnHisGlnCysVal 345
Db 103 AANAATGCTTAACATGATGCTGATGGATTATTACGTGTAGACCCCTAAACATGTATCGGTT 162
QY 346 ---LysLysGlnCysProGluAsnSerGlyCysPheArgHisLeuAspGluArgGluGlu 364
Db 163 GATACAAGAGATATCTCTACAAATCTGGATGTTTTAGAGATCATATAATGCTACTGAAGAA 222
QY 365 CysLysCysLeuLeuAsnThrLysGln---GluGlyAspLysCysValGluAsnProAsn 383
Db 223 TGGAGATGTTTATTACGTTACAAACAAGGTGAAGGTAATACATGTTGTAGAAAATAATAAT 282
QY 384 ProThrCysAsnGluAsnAsnGlyLysCysAspAlaAspAlaThr 398
Db 283 CCTACTTGATATCAACAATGCTGGATGTGATCCAACTGCTACT 327

RESULT 33
BI094571
LOCUS
DEFINITION
LbSSH00002 Preinfection stage symbiosis-regulated cDNAs from L.
bicolor x P. resinosa Laccaria bicolor cDNA, mRNA sequence.
ACCESSION
BI094571
VERSION
BI094571.1 GI:14516049
SOURCE
Laccaria bicolor.
ORGANISM
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
Agaricales; Tricholomataceae; Laccaria.
REFERENCE
Podila,G.K., Brand,J.R. and Hynes,M.J.
Preinfection stage symbiosis-regulated cDNAs from L. bicolor x P.
resinosa
JOURNAL
Unpublished (2001)
CONTACT: Dr. G.K. Podila
Department of Biological Sciences
University of Alabama
Huntsville, AL 35899, USA
Tel: 256 824 6263
Fax: 256 824 6305
Email: podilag@email.uah.edu.
FEATURES
Location/Qualifiers
1..624
/organism="Laccaria bicolor"
/db_xref="taxon:29883"
/clone_lib="Preinfection stage symbiosis-regulated cDNAs
from L. bicolor x P. resinosa"
/note="cDNA clones were selected using a combination of
Suppressive subtraction hybridization derived probes and
a cDNA library of L. bicolor, prepared from pooled RNA of
interaction time points ranging from 6-72 h. Differential
expression of these cDNAs was confirmed through screening
a membrane array of the cDNAs with exponential probes
prepared from control RNA obtained from free living L.

```

bicolor and pooled interaction RNA."

```

BASE COUNT      192 a   124 c   156 g   151 t   1 others
ORIGIN

Alignment Scores:
Pred. No.:      8.79e-11      Length:      624
Score:          192.50      Matches:      59
Percent Similarity: 58.14%      Conservative: 16
Best Local Similarity: 45.74%      Mismatches:  37
Query Match:      8.46%      Indels:       18
DB:              13      Gaps:         3

US-10-057-531A-2 (1-431) x BI094571 (1-624)

QY 3 HisHisHisHisSerSerGlyLeuVal-ProArgLysSerGlyMetLysGluThrAl 22
Db 228 AATCATCAATCATCTCTCTGCTGCTCCACCGCGTTCGTATGAAGAAACCGC 287
QY 22 aAlaAlaLysPheGluArg-GlnHisMetAspSerPro----- 34
Db 288 TGCTGCTAAATTCGACGCCACACATGCACAGCCAGATCTGCATTTAGGTGACACTA 347
QY 35 -----AspLeuGlyThrAspAspAspLysAlaMetAlaAspIleGlySerI 51
Db 348 TAGAATACCAAGATCTGGGTACCGACGACGACGACAGCCATGCGC-GATATCGATCC- 405
QY 51 leGluGlyArgGlyThrMetAlaIleSerValThrMetAspAsnIleLeuSerGlyPheG 71
Db 406 -----GAATTCGGCGCGCTCGATACGCTGCAGAGACGACGAGGCGGGAAGCAG 460
QY 71 luAsnGluTyArgValIleTyLeuLys---ProLeuAlaGlyValTyArgSerLeuL 90
Db 461 CAAGCAGGGTAAACGTGATTTGGATGAACCTTGAGGCACGATGATCTATCAGGATTTGGA 520
QY 90 yLysGlnIleGluLysAsnIlePheThrPheAsnLeu-AsnLeuAsnAspIleLeuAsn 109
Db 521 TGAGCTTGACTGAGCGGGATTATTCGTTATATGATATGATCTAGTAGGAAGATAATAA 580

RESULT 34
BI168979
LOCUS
DEFINITION
EST571502 PyBS Plasmodium yoelii yoelii cDNA clone PYCPK24 5' end,
mRNA sequence.
ACCESSION
BI168979
VERSION
BI168979.1 GI:17302211
KEYWORDS
EST.
SOURCE
Plasmodium yoelii yoelii.
ORGANISM
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE
1 (bases 1 to 780)
Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Vaidya,A.B.,
Fraser,C.M. and Carucci,D.J.
Plasmodium yoelii EST project at TIGR
JOURNAL
Unpublished (2001)
CONTACT: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference
Reagent Resource Center, ATCC
http://www.malaria.mr4.org/mr4pages/index.html
Seq primer: ADF.
FEATURES
Location/Qualifiers
1..780
/organism="Plasmodium yoelii yoelii"
/strain="17XL"

```

```

/db_xref="taxon:73239"
/clone_lib="pYCPK24"
/dev_stage="Asexual blood stages"
/lab_host="E. coli XL-1 Blue"
/notes="vector: pAD-GAL4; At 20-25% parasitemia, blood was
collected from BALB/cBYJ mice infected with Pyl17XL
parasites, and leukocytes removed by passage over
microcrystalline cellulose columns. Total RNA was
isolated using the guanidinium isothiocyanate method, and
mRNA isolated using oligo(dT)-cellulose chromatography.
First strand cDNA synthesis was completed using a 50-base
primer and reverse transcriptase in the presence of
5-methyl dCTP. After second strand synthesis, uneven
termini were treated with pfu DNA polymerase and EcoRI
adaptors ligated to the blunt ends. The sample was cleaved
with XhoI and separated on a Sephacryl S-500 column.
Size-fractionated cDNA was precipitated and ligated to
HybridZAP arms directionally using EcoRI-XhoI cleaved arms.
After packaging, the phagemid vector (pAD-GAL4) was
excised from the HybridZAP vector and plasmid DNA
isolated."

```

BASE COUNT 333 a 76 c 109 g 262 t  
ORIGIN

Alignment Scores:  
Pred. No.: 1.98e-10 Length: 780  
Score: 190.50 Matches: 68  
Percent Similarity: 37.15% Conservativeness: 39  
Best Local Similarity: 23.61% Mismatches: 86  
Query Match: 8.37% Indels: 96  
DB: 13 Gaps: 10

US-10-057-531A-2 (1-431) x BM168979 (1-780)

```

Qy 150 GlnLysAsnThrLeuLeuLysSerTyrLysTyrLysLysGluSerValGluAsnAspLeu 169
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 CAGAACCAAGTAATGATTAATACACAAAATGACAAACAAATGACTACT-AATGATGTA 59
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 170 LysPheAlaGlnGluGlyLeuTyrTyrGluLysValLeuAlaLysTyrLysAspAsp 189
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 60 AAAGATGAAGAAGAGATATTCCTGTAGTAATGGGATACTACTCTAAGTTTAAATAAT 119
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 190 LeuGluSerIleLysValLysGluGlyLysGluLysGluLysPheProSerSerPro 209
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 120 AATGAACCTCAATACTACTATTGAAATAATCCAGAAAAA----- 161
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 210 ThrThrProSerProAlaLysThrAspGluGlnLysLysGluSerLysPheLeuPro 229
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 162 -----AAAAAGAAATAAAA----- 176
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 230 PheLeuThrAsnIleGluThrLeuTyrAsnAsnLeuValAsnLysIleAspAspTyrLeu 249
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 177 -----GAAGATATAATAGATCATGCTGATATAATGCGT 209
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 250 IleAsnLeuLysAlaLys-----IleAsnAspCysAsnValGluLysAspGluAla 266
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 210 ATAACATTGTTAGCAGAAATAGAAATCTTAAAGAAACAACTAGGATAGTAGAAGCA 269
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 267 HisValLysIleThrLysSerAspLysAlaIleAspLysIleAspLysLeuPhe 286
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 270 GTATATAAGATTGAAGTTT-----GTTTAA 296
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 287 LysAsnProTyrAspPheGluAlaIleLysLysLeuIleAsnAspThrLysLysAsp 306
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 297 AAATGTAATTTTATATACTGCTATTACTAATTTTGTCT----- 335
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 307 MetLeuGlyLysLeuLeuSerThrGlyLeuValGlnAsnPheProAsnThrIleIleSer 326
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 335 ----- 335
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 327 LysLeuIleGluGlyLysPheGlnAspMetLeuAsnIleSerGlnHisGlnCysValLys 346
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 336 -----AAATTTAAA-----ACTAATCATATATATGCGAATAT 365
Qy 347 LysGlnCysProGluAsnSerGlyCysPheArgHisLeuAspGluArgGluGluCysLys 366
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 366 TCGAAATGTGTGACAAATGACGATGTTATATTATTGAGAAGATAGCAAGAATGTCGA 425
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 367 CysLeuLeuAsnTyrLysGlnGluGlyAsp-----LysCysValGluAsnPro 382
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 426 TGTTTAGCTAATTATGTACGAGATAGTGTAGATAATATTTTAAATGTGTCCATGCT 485
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 383 AsnProThrCysAsnGluAsnAsnGlyCysAspAlaAspAlaThrCysThrGluGlu 402
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 486 ATTAAGATTGTAATAATAATGGAATTTGTAATAAAATGCTGAATGCTCAATTTAA 545
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 403 AspSerGlySerArgLysLysIleThrCysGluCysThrLysProAspSerTyrPro 422
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 546 AATTAAT-----AAATTTATATGTCATGCT-----TCATATGAT 578
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 423 LeuPhe---AspGlyIlePheCys 429
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 579 TATTTTGGAGATGATTTTGTG 602
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 35
BI094596 693 bp mRNA linear EST 20-JUN-2001
DEFINITION LBSSH00027 Preinfection stage symbiosis-regulated cDNAs from L.
bicolor x P. resinosa Laccaria bicolor cDNA, mRNA sequence.
ACCESSION BI094596
VERSION BI094596.1 GI:14516074
KEYWORDS EST.
SOURCE Laccaria bicolor.
ORGANISM Laccaria bicolor.
REFERENCE Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
AUTHORS Agaricales; Tricholomataceae; Laccaria.
TITLE Podila,G.K., Brand,J.R. and Hynes,M.J.
JOURNAL Preinfection stage symbiosis-regulated cDNAs from L. bicolor x P.
COMMENT resinosae
Contact: Dr. G.K. Podila
Department of Biological Sciences
University of Alabama
Huntsville, AL 35899, USA
Tel: 256 824 6263
Fax: 256 824 6305
Email: podilag@mail.uah.edu.
FEATURES
source Location/Qualifiers
1..693
/organism="Laccaria bicolor"
/db_xref="taxon:29883"
/clone_lib="Preinfection stage symbiosis-regulated cDNAs
from L. bicolor x P. resinosa"
/note="cDNA clones were selected using a combination of
Suppressive subtraction hybridization derived probes and
a cDNA library of L. bicolor, prepared from pooled RNA of
interaction time points ranging from 6-72h. Differential
expression of these cDNAs was confirmed through screening
a membrane array of the cDNAs with exponential probes
prepared from control RNA obtained from free living L.
bicolor and pooled interaction RNA."
BASE COUNT 189 a 164 c 153 g 187 t
ORIGIN

Alignment Scores:
Pred. No.: 2.17e-10 Length: 693
Score: 189.50 Matches: 49
Percent Similarity: 75.38% Conservativeness: 0
Best Local Similarity: 75.38% Mismatches: 1
Query Match: 8.33% Indels: 16
DB: 13 Gaps: 1

US-10-057-531A-2 (1-431) x BI094596 (1-693)

```

QY 1 MethHisHsHisHsHisSerSerGlyLeuValProArg-GlySerGlyMetLysG1 20  
 Db 193 ATGCACCATCATCATCATCTCTTCTGGTCTGGTCCACGCGGTTCTGTTATGAAGA 252  
 QY 20 uThr-AlaAlaAlaLysPheGluArgGln-HisMetAspSer----- 33  
 Db 253 AACCGGCTGCTGCTAAATTCGAACGCCAACACATGACGACGCCCCAGATCTGCATTTAGG 312  
 QY 34 -----Pro-AspLeuGlyThrAspAspAspLysAlaMetAlaAspI1 48  
 Db 313 TGACACTATAGATACCAAGATCTGGTACCGACGACGACAGCCCATGCC-GATAT 371  
 QY 48 eGlySer 50  
 Db 372 CGGATCC 378

RESULT 36  
 BI094595 608 bp mRNA linear EST 20-JUN-2001  
 LOCUS LBSH00026 Preinfection stage symbiosis-regulated cDNAs from L.  
 DEFINITION bicolor x P. resinosa Laccaria bicolor cDNA, mRNA sequence.  
 ACCESSION BI094595  
 VERSION BI094595.1 GI:14516073  
 KEYWORDS EST.  
 SOURCE Laccaria bicolor.  
 ORGANISM Laccaria bicolor.  
 Agaricales; Tricholomataceae; Laccaria.  
 Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;  
 1 (bases 1 to 608)  
 Podila, G.K., Brand, J.R. and Hynes, M.J.  
 AUTHORS Preinfection stage symbiosis-regulated cDNAs from L. bicolor x P.  
 TITLE resinosa  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Dr. G.K. Podila  
 Department of Biological Sciences  
 University of Alabama  
 Huntsville, AL 35899, USA  
 Tel: 256 824 6263  
 Fax: 256 824 6305  
 Email: podila@uah.edu.

FEATURES  
 source  
 1..608  
 Location/Qualifiers  
 /organism="Laccaria bicolor"  
 /db\_xref="taxon:29883"  
 /clone\_lib="Preinfection stage symbiosis-regulated cDNAs  
 from L. bicolor x P. resinosa"  
 /note="cDNA clones were selected using a combination of  
 suppressive subtraction hybridization derived probes and  
 a cDNA library of L. bicolor, prepared from pooled RNA of  
 interaction time points ranging from 6-72 h. Differential  
 expression of these cDNAs was confirmed through screening  
 a membrane array of the cDNAs with exponential probes  
 prepared from control RNA obtained from free living L.  
 bicolor and pooled interaction RNA."  
 BASE COUNT 172 a 141 c 167 g 128 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 2,66e-10 Length: 608  
 Score: 189.00 Matches: 47  
 Percent Similarity: 77.05% Conservativeness: 0  
 Best Local Similarity: 77.05% Mismatches: 3  
 Query Match: 8.26% Indels: 14  
 DB: 13 Gaps: 1

US-10-057-531A-2 (1-431) x BI094595 (1-608)

QY 1 MethHisHsHisHsHisSerSerGlyLeuValProArgGlySerGlyMetLysGlu 20  
 Db 341 ATGCACCATCATCATCATCTCTTCTGGTCTGGTCCACGCGGTTCTGTTATGAAGA-GAA 399  
 QY 21 ThrAlaAlaAlaLysPheGluArgGlnHisMetAspSerProAspLeu----- 36  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 400 ACCGCTGCTGCTAAATTCGAACGCCACACATGACGACGCCAGATCT-GCATTTAGGTGA 458  
 QY 37 -----GlyThrAspAspAspLysAlaMetAlaAspIleGly 49  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 459 CACTATAGATACCAAGTCTGGGTACCGACGACGACCAAGCCATGGC-GATATCGGA 517  
 QY 50 Ser 50  
 ||||  
 Db 518 TCC 520

RESULT 37  
 BM166573 652 bp mRNA linear EST 04-DEC-2001  
 LOCUS EST569096 PyBS Plasmodium yoelii cDNA clone pYCNT91 5' end,  
 DEFINITION mRNA sequence.  
 ACCESSION BM166573  
 VERSION BM166573.1 GI:17299805  
 KEYWORDS EST.  
 SOURCE Plasmodium yoelii yoelii.  
 ORGANISM Plasmodium yoelii yoelii.  
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 REFERENCE 1 (bases 1 to 652)  
 AUTHORS Carlton, J.M., Daly, T.M., Long, C.A., Bergman, L.W., Vaidya, A.B.,  
 Fraser, C.M. and Carucci, D.J.  
 TITLE Plasmodium yoelii EST project at TIGR  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Jane Carlton  
 Parasite Genomics Group  
 The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-530-9319  
 Fax: 301-838-0208  
 Email: carlton@tigr.org  
 For clone info, please contact the Malaria Research and Reference  
 Reagent Resource Center, ATCC  
 http://www.malaria.mr4.org/mr4pages/Index.html  
 Seq primer: ADF.

FEATURES  
 source  
 1..652  
 Location/Qualifiers  
 /organism="Plasmodium yoelii yoelii"  
 /strain="17XL"  
 /db\_xref="taxon:73239"  
 /clone\_lib="PyBS"  
 /dev\_stage="Asexual blood stages"  
 /lab\_host="E. coli XL-1 Blue"  
 /note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was  
 collected from BALB/cByJ mice infected with Py17XL  
 parasites, and leukocytes removed by passage over  
 microcrystalline cellulose columns. Total RNA was  
 isolated using the guanidinium isothiocyanate method, and  
 mRNA isolated using oligo(dT)-cellulose chromatography.  
 First strand cDNA synthesis was completed using a 50-base  
 primer and reverse transcriptase in the presence of  
 5-methyl dCTP. After second strand synthesis, uneven  
 termini were treated with pfu DNA polymerase, and EcoRI  
 adaptors ligated to the blunt ends. The sample was cleaved  
 with XhoI and separated on a Sephacryl S-500 column.  
 Size-fractionated cDNA was precipitated and ligated to  
 HybriZAP arms directionally using EcoRI-XhoI cleaved arms.  
 After packaging, the phagemid vector (PAD-GAL4) was  
 excised from the HybriZAP vector and plasmid DNA  
 isolated."  
 BASE COUNT 290 a 97 c 117 g 148 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 8,05e-10 Length: 652  
 Score: 184.00 Matches: 51  
 Percent Similarity: 50.31% Conservativeness: 29  
 Best Local Similarity: 32.08% Mismatches: 63  
 Query Match: 8.09% Indels: 16  
 DB: 13 Gaps: 4



into EcoRI and XhoI sites of 1 zapII vector using the Zap cDNA synthesis kit (Stratagene, CA). The average size of the cDNA inserts in the library was between 1.0 and 1.5kb. Clones were mass excised using the ExAssist helper phage (Stratagene), the phagemids were precipitated with PEG 8000 and extracted with phenol/chloroform. Phagemid DNA was electroporated into DH10B cells."

BASE COUNT 153 a 32 c 33 g 144 t  
ORIGIN

Alignment Scores:  
Pred. No.: 3,47e-08 Length: 362  
Score: 166.00 Matches: 30  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 7.30% Indels: 0  
DB: 14 Gaps: 0

US-10-057-531A-2 (1-431) x B0451205 (1-362)

QY 402 GluAspSerGlySerArgLysLysIleThrCysGluCysThrLysProAspSerTyr 421  
|||||  
Db 2 GAGATTCAGGTAGCAGCAAGAAATCACATGTGATGTAACCTGATCTTAT 61  
QY 422 ProLeuPheAspGlyIlePheCysSerSer 431  
|||||  
Db 62 CCACCTTTTCGATGATGTTTCTGCAGTTC 91

RESULT 40  
BM167595  
LOCUS  
DEFINITION  
EST570118 PyBS Plasmodium yoelii cDNA clone PYCON35 5' end,  
mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Plasmodium yoelii yoelii.  
Plasmodium yoelii yoelii  
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

REFERENCE  
AUTHORS  
Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Vaidya,A.B.,  
Fraser,C.M. and Carucci,D.J.  
Plasmodium yoelii EST project at TIGR  
Unpublished (2001)  
Contact: Jane Carlton  
Parasite Genomics Group

The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-530-9319  
Fax: 301-838-0208  
Email: carlton@tigr.org

For clone info, please contact the Malaria Research and Reference  
Reagent Resource Center, ATCC  
http://www.malaria.mr4.org/mr4pages/index.html  
Seq primer: ADF.

FEATURES  
source  
1. .686  
/organism="Plasmodium yoelii yoelii"  
/strain="17XL"  
/db\_xref="taxon:73239"  
/clone="PYCON35"  
/clone\_lib="PyBS"  
/dev\_stage="Asexual blood stages"  
/lab\_host="E. coli XL-1 Blue"  
/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was collected from BALB/cByJ mice infected with Py17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidinium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven

termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybrizAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (PAD-GAL4) was excised from the HybrizAP vector and plasmid DNA isolated."

BASE COUNT 283 a 108 c 130 g 165 t  
ORIGIN

Alignment Scores:  
Pred. No.: 1.37e-07 Length: 686  
Score: 164.00 Matches: 47  
Percent Similarity: 50.00% Conservative: 23  
Best Local Similarity: 33.57% Mismatches: 54  
Query Match: 7.21% Indels: 16  
DB: 13 Gaps: 4

US-10-057-531A-2 (1-431) x BM167595 (1-686)

QY 10 GlyLeuValProArgGlySerGlyMetLysGluThrAlaAlaLysPheGluArgGln 29  
|||||  
Db 315 GGTGCAGTACCA---GGATCAGGAACCGATACACGGGTAGCTGGA----- 356  
QY 30 HisMetAspSerProAspLeuGlyThrAspAspLysAlaMetAlaAspIleGly 49  
|||||  
Db 357 -----AGCAGTGTGATGATAATGAAGACGATCATATATCAAAATGCAAGT---GGT 407

QY 50 SerIleGluGlyArgGlyThrMetAlaIleSerValThrMetAspAsnIleLeuSerGly 69  
|||||  
Db 408 CAATCCGAA-----GATGCACCCAGAAAAGATATTTCTTCGAA 446

QY 70 PheGluAsnGluTyrAspValIleTyrLeuLysProLeuAlaGlyValTyrArgSerLeu 89  
|||||  
Db 447 TTTACAAATGAAAGTTTGTATGATATACAAAAGGTTGGGTAGTACATATAATCATTA 506

QY 90 LysLysGlnIleGluLysAsnIlePheThrPheAsnLeuAsnLeuAsnAspIleLeuAsn 109  
|||||  
Db 507 AAGAAACACATGTTAAGAGAAATTTTCAACAATTAAGACACATGACAAATGGATTAAAT 566

QY 110 SerArgLeuLysLysArgLysTyrPheLeuAspValLeuGluSerAspLeuMetGlnPhe 129  
|||||  
Db 567 AATAAATCACAATAAGAAATGATTTCTTGAAGTATTAAAGCCCATGAATTAGATTATTC 626

QY 130 LysHisIleSerSerAsnGluTyrIleIleGluAspSerPheLysLeuLeuAsnSerGlu 149  
|||||  
Db 627 AAAGATTAAAGTACCAACAATATGTTATTAGAAATCCATATCAATTATTAGATGATGAT 686

Search completed: May 19, 2003, 16:14:54  
Job time : 1292 secs





GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 10:19:28; Search time 51 Seconds  
(without alignments)  
777.707 Million cell updates/sec

Title: US-10-057-531A-2

Perfect score: 2275

Sequence: 1 MHHHHHSSGLVPRSGMKE.....TCECTKPSYFLPDGIFCSS 431

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 349150 seqs, 92025710 residues

Total number of hits satisfying chosen parameters: 349150

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA:\*

1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*

2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*

3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*

4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*

5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*

6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*

7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*

8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*

9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*

10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*

11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*

12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*

13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*

14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1965	86.4	394	9	US-09-978-756-3
2	1928	84.7	402	9	US-10-098-514-14
3	1911	84.0	383	9	US-10-098-514-4
4	1900	83.5	383	9	US-10-098-514-2
5	1175	51.6	1639	9	US-10-087-464-10
6	1165.5	51.2	376	9	US-09-978-756-2
7	1165.5	51.2	376	9	US-10-087-464-34
8	1165.5	51.2	378	9	US-10-087-464-11
9	788	34.6	380	10	US-09-134-333-12
10	777	34.2	380	10	US-09-134-333-13
11	734.5	32.3	379	10	US-09-134-333-11
12	527	23.2	108	10	US-09-134-333-10
13	527	23.2	116	10	US-09-134-333-5
14	524	23.0	96	9	US-09-978-756-1
15	524	23.0	114	9	US-10-087-464-35
16	523	23.0	95	10	US-09-134-333-2
17	523	23.0	127	10	US-09-134-333-8
18	424.5	18.7	281	10	US-09-134-333-14
19	306.5	13.5	516	10	US-09-804-626-4

20	301.5	13.3	518	10	US-09-804-626-2	Sequence 2, Appli
21	278.5	12.2	824	9	US-10-046-583A-3	Sequence 3, Appli
22	262.5	11.5	355	9	US-10-012-896-1011	Sequence 1011, Ap
23	258.5	11.4	201	9	US-09-681-938-1	Sequence 1, Appli
24	167	7.3	55	9	US-10-042-945-28	Sequence 28, Appl
25	167	7.3	55	10	US-09-757-417-28	Sequence 25, Appl
26	145.5	6.4	1394	9	US-09-842-930A-25	Sequence 5251, Ap
27	144	6.3	996	10	US-09-815-242-5251	Sequence 12141, A
28	144	6.3	1009	10	US-09-815-242-12141	Sequence 65, Appl
29	140.5	6.2	980	10	US-09-888-615-65	Sequence 21, Appl
30	131.5	5.8	948	9	US-10-267-311-21	Sequence 73, Appl
31	130	5.7	1805	9	US-09-820-843A-73	Sequence 2239, Ap
32	128	5.6	391	9	US-10-046-935-2239	Sequence 2239, Ap
33	128	5.6	391	9	US-10-146-502-2239	Sequence 1160, Ap
34	128	5.6	580	10	US-09-764-864-1160	Sequence 61, Appl
35	127.5	5.6	690	9	US-09-298-523B-61	Sequence 62, Appl
36	127.5	5.6	701	9	US-09-298-523B-62	Sequence 7, Appli
37	126	5.5	506	9	US-10-157-223-7	Sequence 63, Appl
38	125	5.5	670	9	US-09-298-523B-63	Sequence 14, Appl
39	123	5.4	589	9	US-09-298-523B-14	Sequence 1, Appli
40	123	5.4	691	9	US-09-298-523B-1	Sequence 60, Appl
41	123	5.4	929	9	US-09-298-523B-60	Sequence 6, Appli
42	122.5	5.4	2710	9	US-10-011-366-6	Sequence 35, Appl
43	121	5.3	344	9	US-09-820-843A-35	Sequence 127, App
44	121	5.3	905	9	US-10-114-893-127	Sequence 59, Appl
45	121	5.3	915	10	US-09-880-192-59	

ALIGNMENTS

RESULT 1

US-09-978-756-3

; Sequence 3, Application US/09978756

; Patent No. US20020160017A1

; GENERAL INFORMATION:

; APPLICANT: Holder, Anthony

; APPLICANT: Birdsall, Berry

; APPLICANT: Feeney, James

; APPLICANT: Morgan, William

; APPLICANT: Syed, Shabih

; TITLE OF INVENTION: Malaria Vaccine

; FILE REFERENCE: 18396/1005

; CURRENT APPLICATION NUMBER: US/09/978,756

; CURRENT FILING DATE: 2001-10-16

; PRIOR APPLICATION NUMBER: PCT/GB00/01558

; PRIOR FILING DATE: 2000-04-20

; PRIOR APPLICATION NUMBER: 09/311,817

; PRIOR FILING DATE: 1999-05-13

; PRIOR APPLICATION NUMBER: 2,271,451

; PRIOR FILING DATE: 1999-05-25

; PRIOR APPLICATION NUMBER: 9909072.2

; PRIOR FILING DATE: 1999-04-20

; NUMBER OF SEQ ID NOS: 3

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 3

; LENGTH: 394

; TYPE: PRT

; ORGANISM: Plasmodium falciparum

US-09-978-756-3

Query Match 86.4%; Score 1965; DB 9; Length 394;

Best Local Similarity 100.0%; Pred. No. 1.8e-117;

Matches 374; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 AISVTMDNLSGGENEYDVYILKPLAGVYRSLKKQIEKNFTNLDNLNLSRLKKRY 117

Db 1 AISVTMDNLSGGENEYDVYILKPLAGVYRSLKKQIEKNFTNLDNLNLSRLKKRY 60

QY 118 FLDVLESDLMOFKHSSNEVYIIEDSFKLLNSEQNTLLKSKYIKESVENDIKFAQEGIS 177

Db 61 FLDVLESDLMOFKHSSNEVYIIEDSFKLLNSEQNTLLKSKYIKESVENDIKFAQEGIS 120

Qy 178 YEKVLAKYKDDLSIKKVIKEEKEKFPSPPTTPPSAKTDEQKESKFLPFLTNIETL 237  
Db 121 YEKVLAKYKDDLSIKKVIKEEKEKFPSPPTTPPSAKTDEQKESKFLPFLTNIETL 180  
Qy 238 YNNLVNKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFPKNPYDFAIKK 297  
Db 181 YNNLVNKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFPKNPYDFAIKK 240  
Qy 298 LINDTKKMDLGLKLLSTGLVQNFNTIISKLEIEGKFQDMLNISQHCYKVKQCPENSGCFR 357  
Db 241 LINDTKKMDLGLKLLSTGLVQNFNTIISKLEIEGKFQDMLNISQHCYKVKQCPENSGCFR 300  
Qy 358 HLDRECKCLLNKYQEGDKCVENPNPTCNENNGCGDADATCTEEDSGSRKKITCECTK 417  
Db 301 HLDRECKCLLNKYQEGDKCVENPNPTCNENNGCGDADATCTEEDSGSRKKITCECTK 360  
Qy 418 PDSYPLFDGIFCSS 431  
Db 361 PDSYPLFDGIFCSS 374

RESULT 2

US-10-098-514-14  
; Sequence 14, Application us/10098514  
; Publication No. US20020194648A1  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Sandra P  
; APPLICANT: Christopher, David A  
; APPLICANT: Vine, Benjamin  
; APPLICANT: Su, Wei-Wen  
; APPLICANT: Bugos, Robert  
; TITLE OF INVENTION: PLASMODIUM FALCIPARUM MEROZOITE SURFACE PROTEIN-1 MALARIA PRODUCE  
; FILE REFERENCE: A-71339/RFT/TAL/NBC  
; CURRENT FILING DATE: 2002-08-06  
; PRIOR APPLICATION NUMBER: US 10/098,514  
; PRIOR FILING DATE: 2000-02-08  
; PRIOR APPLICATION NUMBER: US 09/500,376  
; PRIOR FILING DATE: 2000-02-08  
; PRIOR APPLICATION NUMBER: US 60/274,599  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 14  
; LENGTH: 402  
; TYPE: PRT  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: synthetic  
US-10-098-514-14

Query Match 84.7%; Score 1928; DB 9; Length 402;  
Best Local Similarity 98.1%; Pred. No. 4.1e-115;  
Matches 368; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
Qy 56 TMAISVTMDNLTSGFENEYDVIYKPLAGVYRSLKKQIEKNIFTNLDNLINSLRKKR 115  
Db 20 TAAISVTMDNLTSGFENEYDVIYKPLAGVYRSLKKQIEKNIFTNLDNLINSLRKKR 79  
Qy 116 KYFLDVLESQDMQFKHISSNEYIIEDSFKLLNSEQKNTLLSKYKIKESVENDIKFAQEG 175  
Db 80 KYFLDVLESQDMQFKHISSNEYIIEDSFKLLNSEQKNTLLSKYKIKESVENDIKFAQEG 139  
Qy 176 ISYEKVLAKYKDDLSIKKVIKEEKEKFPSPPTTPPSAKTDEQKESKFLPFLTNIE 235  
Db 140 ISYEKVLAKYKDDLSIKKVIKEEKEKFPSPPTTPPSAKTDEQKESKFLPFLTNIE 199  
Qy 236 TLYNNLVNKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFPKNPYDFAIK 295  
Db 200 TLYNNLVNKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFPKNPYDFAIK 259  
Qy 296 KKLINDTKKMDLGLKLLSTGLVQNFNTIISKLEIEGKFQDMLNISQHCYKVKQCPENSGC 355  
Db 260 KKLINDTKKMDLGLKLLSTGLVQNFNTIISKLEIEGKFQDMLNISQHCYKVKQCPENSGC 319

Qy 356 FRHLDRECKCLLNKYQEGDKCVENPNPTCNENNGCGDADATCTEEDSGSRKKITCEC 415  
Db 320 FRHLDRECKCLLNKYQEGDKCVENPNPTCNENNGCGDADATCTEEDSGSGNGKTIICEC 379  
Qy 416 TKPDSYPLFDGIFCS 430  
Db 380 TKPDSYPLFDGIFCS 394

RESULT 3

US-10-098-514-4  
; Sequence 4, Application us/10098514  
; Publication No. US20020194648A1  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Sandra P  
; APPLICANT: Christopher, David A  
; APPLICANT: Vine, Benjamin  
; APPLICANT: Su, Wei-Wen  
; APPLICANT: Bugos, Robert  
; TITLE OF INVENTION: PLASMODIUM FALCIPARUM MEROZOITE SURFACE PROTEIN-1 MALARIA PROD  
; FILE REFERENCE: A-71339/RFT/TAL/NBC  
; CURRENT FILING DATE: 2002-08-06  
; PRIOR APPLICATION NUMBER: US 10/098,514  
; PRIOR FILING DATE: 2000-02-08  
; PRIOR APPLICATION NUMBER: US 09/500,376  
; PRIOR FILING DATE: 2000-02-08  
; PRIOR APPLICATION NUMBER: US 60/274,599  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 383  
; TYPE: PRT  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: synthetic  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (380)..(380)  
; OTHER INFORMATION: "Xaa" at position 380 represents a stop codon  
US-10-098-514-4

Query Match 84.0%; Score 1911; DB 9; Length 383;  
Best Local Similarity 98.1%; Pred. No. 4.7e-114;  
Matches 363; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 61 VTMDNLTSGFENEYDVIYKPLAGVYRSLKKQIEKNIFTNLDNLINSLRKKRYFLD 120  
Db 6 ITMDNLTSGFENEYDVIYKPLAGVYRSLKKQIEKNIFTNLDNLINSLRKKRYFLD 65  
Qy 121 VLESQDMQFKHISSNEYIIEDSFKLLNSEQKNTLLSKYKIKESVENDIKFAQEGISYVE 180  
Db 66 VLESQDMQFKHISSNEYIIEDSFKLLNSEQKNTLLSKYKIKESVENDIKFAQEGISYVE 125  
Qy 181 KVLAKYKDDLSIKKVIKEEKEKFPSPPTTPPSAKTDEQKESKFLPFLTNIETLYNN 240  
Db 126 KVLAKYKDDLSIKKVIKEEKEKFPSPPTTPPSAKTDEQKESKFLPFLTNIETLYNN 185  
Qy 241 LVNKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFPKNPYDFAIKKLN 300  
Db 186 LVNKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFPKNHNDFAIKKLN 245  
Qy 301 DDTKKMDLGLKLLSTGLVQNFNTIISKLEIEGKFQDMLNISQHCYKVKQCPENSGCFRHL 360  
Db 246 DDTKKMDLGLKLLSTGLVQNFNTIISKLEIEGKFQDMLNISQHCYKVKQCPENSGCFRHL 305  
Qy 361 ERECKCLLNKYQEGDKCVENPNPTCNENNGCGDADATCTEEDSGSRKKITCECTKPD 420  
Db 306 ERECKCLLNKYQEGDKCVENPNPTCNENNGCGDADATCTEEDSGSGNGKTIICECTKPD 365  
Qy 421 YPLFDGIFCS 430  
Db 421 YPLFDGIFCS 430

Db 366 YPLFDGIFCS 375

RESULT 4

US-10-098-514-2  
; Sequence 2, Application US/10098514  
; Publication No. US20020194648A1  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Sandra P  
; APPLICANT: Christopher, David A  
; APPLICANT: Vine, Benjamin  
; APPLICANT: Su, Wei-Wen  
; APPLICANT: Buge, Robert  
; TITLE OF INVENTION: PLASMODIUM FALCIPARUM MEROZOITE SURFACE PROTEIN-1 MALARIA PRODUCE  
; FILE REFERENCE: A-71339/RET/TAL/NBC  
; CURRENT APPLICATION NUMBER: US/10/098,514  
; CURRENT FILING DATE: 2002-08-06  
; PRIOR APPLICATION NUMBER: US 09/500,376  
; PRIOR FILING DATE: 2000-02-08  
; PRIOR APPLICATION NUMBER: US 60/274,599  
; PRIOR FILING DATE: 2001-03-09  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 383  
; TYPE: PRT  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: synthetic  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (380)..(380)  
; OTHER INFORMATION: "Xaa" at position 380 represents a stop codon  
US-10-098-514-2

Query Match 83.5%; Score 1900; DB 9; Length 383;

Best Local Similarity 97.3%; Pred. No. 2.3e-113;

Matches 361; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 60 SVTMDNLISGFENEYDVIYKPLAGYVRSLSKQIEKNIFTNLDILNSRLKRRKYFL 119  
Db 5 NILSONILSGFENEYDVIYKPLAGYVRSLSKQIEKNIFTNLDILNSRLKRRKYFL 64  
QY 120 DVLESOLMFKHISSEYIIESFLLNSEQNTLLKSYIKESVENDIKFAQEGISY 179  
Db 65 DVLESOLMFKHISSEYIIESFLLNSEQNTLLKSYIKESVENDIKFAQEGISY 124  
QY 180 EKVLAQYKDDLESIKKVIKEEKFPSPPTTPPSAKTDEQKESKFLPFLTNIELYN 239  
Db 125 EKVLAQYKDDLESIKKVIKEEKFPSPPTTPPSAKTDEQKESKFLPFLTNIELYN 184  
QY 240 NLVNTKIDDLINLKAINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNPYDPEAKKLI 299  
Db 185 NLVNTKIDDLINLKAINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHDEDAKLI 244  
QY 300 NDDTKRDMGLKLLSTGLVGNFPNTIISKIEGKFDMLNISQHCYKQCPENSGCFRHL 359  
Db 245 NDDTKRDMGLKLLSTGLVGNFPNTIISKIEGKFDMLNISQHCYKQCPENSGCFRHL 304  
QY 360 DERECKCLLNYKQEGDKCVENPNPTCNENNGCGDADATCTEEDSGSRKKTCTCTKPD 419  
Db 305 DERECKCLLNYKQEGDKCVENPNPTCNENNGCGDADATCTEEDSGSRKKTCTCTKPD 364  
QY 420 SYPLFDGIFCS 430  
Db 365 SYPLFDGIFCS 375

RESULT 5

US-10-087-464-10  
; Sequence 10, Application US/10087464  
; Publication No. US20030059436A1

GENERAL INFORMATION:

; APPLICANT: Chishti, Athar  
; APPLICANT: Oh, Steven  
; APPLICANT: Liu, David  
; APPLICANT: Goel, Vikas  
; APPLICANT: Li, Xuerong  
; TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses There  
; FILE REFERENCE: S1237/7019  
; CURRENT APPLICATION NUMBER: US/10/087,464  
; CURRENT FILING DATE: 2002-03-01  
; PRIOR APPLICATION NUMBER: US 06/272,930  
; PRIOR FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 10  
; LENGTH: 1639  
; TYPE: PRT  
; ORGANISM: Plasmodium falciparum  
US-10-087-464-10

Query Match 51.6%; Score 1175; DB 9; Length 1639;

Best Local Similarity 56.6%; Pred. No. 1.5e-66;

Matches 226; Conservative 59; Mismatches 90; Indels 24; Gaps 4;

QY 34 PDLGTDDDDKAMADIGSIEGRGTMAISVTMDNLSGFENEYDVIYKPLAGYVRSLSKQI 93  
Db 1244 PIFGESEED--YDDLQGVVTGEAVTPSV-IDNLSKIENEYEVLYLKLPLAGYVRSLSKQI 1300  
QY 94 EKNIFTNLDILNSRLKRRKYFLDVLSDLMQKHISSEYIIESFLLNSEQNT 153  
Db 1301 ENNVMTFNVNVDILNSRFNKNRENFVNLESULPIKYDTSNNVYVVKDPYKFLNKKRDK 1360  
QY 154 LKSYKYIKESVENDIKFAQEGISYKVLAKYKDDLESIKKVIKEEKFPSPPTTPP 213  
Db 1361 FLSSYVINKSDITDINFANDVLGYKILSEKYKSDLSIKKI----- 1404  
QY 214 SPKATDEQKESKFLPFLTNIELYNLNNKIDDLINLKAINDCNVEKDEAHVKITKL 273  
Db 1405 ---NDKQGENEKYLPFLNLTETVNDKIDLFVHLEAKVNLVNYEKSINVEVKIKEL 1460  
QY 274 SLDKAIIDDKIDLFKNPYDPEAKKIDDKIDLFKNPYDPEAKKIDDKIDLFKNPYDPEAKK 332  
Db 1461 NYLKTIDKLDADFKKNNFVGIADLSTDYNNHNLTKFLSTGMVFENLAKTVLSNLDGN 1520  
QY 333 FQDMLNISQHCYKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGG 392  
Db 1521 LQGLMLNISQHCYKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGG 1580  
QY 393 CDADATCTEEDSGSRKKTCTCTKPDSPFLPDGIFCSS 431  
Db 1581 CDADAKTEEDSGSNGKKTCTCTKPDSPFLPDGIFCSS 1619

RESULT 6

US-09-978-756-2

; Sequence 2, Application US/09978756

; Patent No. US20020160017A1

GENERAL INFORMATION:

; APPLICANT: Holder, Anthony  
; APPLICANT: Birdsall, Berry  
; APPLICANT: Feeney, James  
; APPLICANT: Morgan, William  
; APPLICANT: Syed, Shabih  
; TITLE OF INVENTION: Malaria Vaccine  
; FILE REFERENCE: 18396/1005  
; CURRENT APPLICATION NUMBER: US/09/978,756  
; CURRENT FILING DATE: 2001-10-16  
; PRIOR APPLICATION NUMBER: PCT/GB00/01558  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: 09/311,817  
; PRIOR FILING DATE: 1999-05-13  
; PRIOR APPLICATION NUMBER: 2,271,451  
; PRIOR FILING DATE: 1999-05-25





RESULT 11  
US-09-134-333-11  
; Sequence 11, Application US/09134333  
; Patent No. US20020076403A1  
; GENERAL INFORMATION:  
; APPLICANT: LONGACRE-ANDRE, SHIRLEY  
; APPLICANT: ROTH, CHARLES  
; APPLICANT: NATO, FARIDABANO  
; APPLICANT: BARNWELL, JOHN  
; APPLICANT: MENDIS, KAMINI  
; TITLE OF INVENTION: RECOMBINANT PROTEIN CONTAINING A C-TERMINAL FRAGMENT OF  
; FILE REFERENCE: 0660-0135-0XCIP  
; CURRENT APPLICATION NUMBER: US/09/134,333  
; EARLIER FILING DATE: 1999-04-18  
; EARLIER FILING DATE: 1997-02-14  
; EARLIER FILING DATE: 1997-02-14  
; EARLIER FILING DATE: 1996-02-14  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 11  
; LENGTH: 379  
; TYPE: PRT  
; ORGANISM: Plasmodium cynomolgi  
; FEATURE:  
; OTHER INFORMATION: Amino Acids 1-139- REGION I  
; FEATURE:  
; OTHER INFORMATION: Amino Acids 140-177-REGION II  
; FEATURE:  
; OTHER INFORMATION: Amino Acids 178-282-REGION III  
; FEATURE:  
; OTHER INFORMATION: Amino Acids 283-379-REGION IV  
US-09-134-333-11

Query Match 32.3%; Score 734.5; DB 10; Length 379;  
Best Local Similarity 41.9%; Pred. No. 2.4e-39;  
Matches 156; Conservative 74; Mismatches 115; Indels 27; Gaps 10;  
Qy 72 NEYDVYLLKPLAGVVRSLKQKIEKNIFTFNLNDILNSRLKRYFLDVLDESDLMOFKH 131  
Db 22 NEYDVYIKPLAGWYKTKIKPLEHVNALNTIIDLMSRLKKNRYFLDVLNSDLNPYSI 81  
Qy 132 ISSNEYIIEFSLNSQKNTLKSYYKESVENDIKFAQEGISYYEKVLAKYKDDLE 191  
Db 82 PHSGEYIKDPYKLLDEKKK-LLGSYKYGASVDKDMVTANDGLAYYQKMGDLYKKHLD 140  
Qy 192 SIKKVIK-----EKEFPSPPTTPSPAKTDEQKE-SKFLPFLNIETLYNNL 241  
Db 141 EVNACIKVEANINKHDEEIKKIGSEASKANDKNQNAKKELOKYLPLSLSSIOKEYSTL 200  
Qy 242 VNKIDDDYLNILKAKINDCNVEKDEAHVKITKLSDLKAIDDDILFKNPYDFEAIKKLND 301  
Db 201 VNKVHSYTDTLKKIINNCOIEKETETIVNKLSDYKDELDVYKQS-----KK--ED 252  
Qy 302 DTKKD-MLGKLLSTGLV-QNFPNTIISKLEKGFQDMLNI-SHQCVKQKQCPENSGCFRH 358  
Db 253 DVKSSGLEKMLNSKLINQESKALSSELLNVQIQ-MLMSSEHRCIDTNPVNAACVRY 311  
Qy 359 LDRECKCLLNYKQEGDKCVENPNPCNNENNGGCDATCTEDSGSRKKITCECTKP 418  
Db 312 LDGTEWRCLLYFKEDAGKCVAPNMTCKDKNGGCAPEACKMND----KNEIVCKCTKE 367  
Qy 419 DSYPLFGIFCS 430  
Db 368 GSEPLFEGVFC 379  
RESULT 12  
US-09-134-333-10  
; Sequence 10, Application US/09134333  
; Patent No. US20020076403A1

; GENERAL INFORMATION:  
; APPLICANT: LONGACRE-ANDRE, SHIRLEY  
; APPLICANT: ROTH, CHARLES  
; APPLICANT: NATO, FARIDABANO  
; APPLICANT: BARNWELL, JOHN  
; APPLICANT: MENDIS, KAMINI  
; TITLE OF INVENTION: RECOMBINANT PROTEIN CONTAINING A C-TERMINAL FRAGMENT OF  
; FILE REFERENCE: 0660-0135-0XCIP  
; CURRENT APPLICATION NUMBER: US/09/134,333  
; EARLIER FILING DATE: 1999-04-18  
; EARLIER FILING DATE: 1997-02-14  
; EARLIER FILING DATE: 1997-02-14  
; EARLIER FILING DATE: 1996-02-14  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 10  
; LENGTH: 108  
; TYPE: PRT  
; ORGANISM: Plasmodium falciparum  
US-09-134-333-10  
Query Match 23.2%; Score 527; DB 10; Length 108;  
Best Local Similarity 94.8%; Pred. No. 8.3e-27;  
Matches 91; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
Qy 335 DMLNISQHCYKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNPCNNENGGCD 394  
Db 13 DEFNISQHCYKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNPCNNENGGCD 72  
Qy 395 ADATCTEEDSGSRKKITCECTKPDSPYPLFDGIFCS 430  
Db 73 ADAKTEEDSGNGKKITCECTKPDSPYPLFDGIFCS 108

RESULT 13  
US-09-134-333-5  
; Sequence 5, Application US/09134333  
; Patent No. US20020076403A1  
; GENERAL INFORMATION:  
; APPLICANT: LONGACRE-ANDRE, SHIRLEY  
; APPLICANT: ROTH, CHARLES  
; APPLICANT: NATO, FARIDABANO  
; APPLICANT: BARNWELL, JOHN  
; APPLICANT: MENDIS, KAMINI  
; TITLE OF INVENTION: RECOMBINANT PROTEIN CONTAINING A C-TERMINAL FRAGMENT OF  
; FILE REFERENCE: 0660-0135-0XCIP  
; CURRENT APPLICATION NUMBER: US/09/134,333  
; EARLIER FILING DATE: 1999-04-18  
; EARLIER FILING DATE: 1997-02-14  
; EARLIER FILING DATE: 1997-02-14  
; EARLIER FILING DATE: 1996-02-14  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 116  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC  
US-09-134-333-5  
Query Match 23.2%; Score 527; DB 10; Length 116;  
Best Local Similarity 96.8%; Pred. No. 9e-27;  
Matches 91; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
Qy 338 NISQHCYKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNPCNNENGGCDADA 397  
Db 3 NISQHCYKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNPCNNENGGCDADA 62

QY 398 TCTEEDSGSRKKITCECTKPDSPYPLFDGIFCSS 431  
Matches 90; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
Db 63 KCTEEDSGSGNGKKITCECTKPDSPYPLFDGIFCSS 96

RESULT 14  
US-09-978-756-1  
; Sequence 1, Application US/09978756  
; Patent No. US20020160017A1  
; GENERAL INFORMATION:  
; APPLICANT: Holder, Anthony  
; APPLICANT: Birdsell, Berry  
; APPLICANT: Feeney, James  
; APPLICANT: Morgan, William  
; APPLICANT: Syed, Shabih  
; TITLE OF INVENTION: Malaria Vaccine  
; FILE REFERENCE: 18396/1005  
; CURRENT APPLICATION NUMBER: US/09/978,756  
; PRIOR FILING DATE: 2001-10-16  
; PRIOR APPLICATION NUMBER: PCT/GB00/01558  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: 09/311,817  
; PRIOR FILING DATE: 1999-05-13  
; PRIOR APPLICATION NUMBER: 2,271,451  
; PRIOR FILING DATE: 1999-05-25  
; PRIOR APPLICATION NUMBER: 9909072.2  
; PRIOR FILING DATE: 1999-04-20  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1  
; LENGTH: 96  
; TYPE: PRT  
; ORGANISM: Plasmodium falciparum  
US-09-978-756-1

Query Match 23.0%; Score 524; DB 9; Length 96;  
Best Local Similarity 95.7%; Pred. No. 1.1e-26;  
Matches 90; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 338 NISOHCQVKKQCPNSGCFRHLDERECKLLNYKQGGDKCVENPNTCNENGGCDADA 397  
Matches 90; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
Db 1 NISOHCQVKKQCPNSGCFRHLDERECKLLNYKQGGDKCVENPNTCNENGGCDADA 60

QY 398 TCTEEDSGSRKKITCECTKPDSPYPLFDGIFCSS 431  
Matches 90; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
Db 61 KCTEEDSGSGNGKKITCECTKPDSPYPLFDGIFCSS 94

RESULT 15  
US-10-087-464-35  
; Sequence 35, Application US/10087464  
; Publication No. US20030059436A1  
; GENERAL INFORMATION:  
; APPLICANT: Chishti, Athar  
; APPLICANT: Oh, Steven  
; APPLICANT: Liu, David  
; APPLICANT: Goel, Vikas  
; APPLICANT: Li, Xuerong  
; TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses Thereof  
; FILE REFERENCE: S1237/7019  
; CURRENT APPLICATION NUMBER: US/10/087,464  
; PRIOR FILING DATE: 2002-03-01  
; PRIOR APPLICATION NUMBER: US 06/272,930  
; PRIOR FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 35  
; LENGTH: 114  
; TYPE: PRT  
; ORGANISM: Plasmodium falciparum  
US-10-087-464-35

Query Match 23.0%; Score 524; DB 9; Length 114;

Best Local Similarity 95.7%; Pred. No. 1.4e-26;  
Matches 90; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
QY 338 NISOHCQVKKQCPNSGCFRHLDERECKLLNYKQGGDKCVENPNTCNENGGCDADA 397  
Matches 90; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
Db 1 NISOHCQVKKQCPNSGCFRHLDERECKLLNYKQGGDKCVENPNTCNENGGCDADA 60

QY 398 TCTEEDSGSRKKITCECTKPDSPYPLFDGIFCSS 431  
Matches 90; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
Db 61 KCTEEDSGSGNGKKITCECTKPDSPYPLFDGIFCSS 94

RESULT 16  
US-09-134-333-2  
; Sequence 2, Application US/09134333  
; Patent No. US20020076403A1  
; GENERAL INFORMATION:  
; APPLICANT: LONGACRE-ANDRE, SHIRLEY  
; APPLICANT: ROTH, CHARLES  
; APPLICANT: NATO, FARIDABANO  
; APPLICANT: BARNWELL, JOHN  
; APPLICANT: MENDIS, KAMINI  
; TITLE OF INVENTION: RECOMBINANT PROTEIN CONTAINING A C-TERMINAL FRAGMENT OF  
; FILE REFERENCE: 0660-0135-0XCIP  
; CURRENT APPLICATION NUMBER: US/09/134,333  
; CURRENT FILING DATE: 1999-04-18  
; EARLIER APPLICATION NUMBER: PCT/FR97/00290  
; EARLIER FILING DATE: 1997-02-14  
; EARLIER APPLICATION NUMBER: FR96/01822  
; EARLIER FILING DATE: 1996-02-14  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 95  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC  
US-09-134-333-2

Query Match 23.0%; Score 523; DB 10; Length 95;  
Best Local Similarity 96.8%; Pred. No. 1.3e-26;  
Matches 90; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 338 NISOHCQVKKQCPNSGCFRHLDERECKLLNYKQGGDKCVENPNTCNENGGCDADA 397  
Matches 90; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
Db 3 NISOHCQVKKQCPNSGCFRHLDERECKLLNYKQGGDKCVENPNTCNENGGCDADA 62

QY 398 TCTEEDSGSRKKITCECTKPDSPYPLFDGIFCSS 430  
Matches 90; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
Db 63 KCTEEDSGSGNGKKITCECTKPDSPYPLFDGIFCSS 95

RESULT 17  
US-09-134-333-8  
; Sequence 8, Application US/09134333  
; Patent No. US20020076403A1  
; GENERAL INFORMATION:  
; APPLICANT: LONGACRE-ANDRE, SHIRLEY  
; APPLICANT: ROTH, CHARLES  
; APPLICANT: NATO, FARIDABANO  
; APPLICANT: BARNWELL, JOHN  
; APPLICANT: MENDIS, KAMINI  
; TITLE OF INVENTION: RECOMBINANT PROTEIN CONTAINING A C-TERMINAL FRAGMENT OF  
; FILE REFERENCE: 0660-0135-0XCIP  
; CURRENT APPLICATION NUMBER: US/09/134,333  
; CURRENT FILING DATE: 1999-04-18  
; EARLIER APPLICATION NUMBER: PCT/FR97/00290  
; EARLIER FILING DATE: 1997-02-14  
; EARLIER APPLICATION NUMBER: FR96/01822  
; EARLIER FILING DATE: 1996-02-14

```

; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 127
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-09-134-333-8

Query Match      23.0%; Score 523; DB 10; Length 127;
Best Local Similarity 96.8%; Pred. No. 1.8e-26;
Matches 90; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 338 NISQHCYKQCPENSGCFRHLDERBECKCLLNKYQEGDKCVENPPTCNENGGCDADA 397
Db 35 NISQHCYKQCPENSGCFRHLDERBECKCLLNKYQEGDKCVENPPTCNENGGCDADA 94

Qy 398 TCTEDSGSRKKITCECTKPDSPFLFDGIFCS 430
Db 95 KCTEDSGSNGKKITCECTKPDSPFLFDGIFCS 127

RESULT 18
US-09-134-333-14
; Sequence 14, Application US/09134333
; Patent No. US20020076403A1
; GENERAL INFORMATION:
; APPLICANT: LONGACRE-ANDRE, SHIRLEY
; APPLICANT: ROTH, CHARLES
; APPLICANT: NATO, FARIDABANO
; APPLICANT: BARNWELL, JOHN
; APPLICANT: MENDIS, KAMINI
; TITLE OF INVENTION: RECOMBINANT PROTEIN CONTAINING A C-TERMINAL FRAGMENT OF
; FILE REFERENCE: 0660-0135-0XCIP
; CURRENT APPLICATION NUMBER: US/09/134,333
; CURRENT FILING DATE: 1999-04-18
; EARLIER APPLICATION NUMBER: PCT/FR97/00290
; EARLIER FILING DATE: 1997-02-14
; EARLIER APPLICATION NUMBER: FR96/01822
; EARLIER FILING DATE: 1996-02-14
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: ALIGNMENT
; FEATURE:
; OTHER INFORMATION: Amino Acids 1-115-REGION I
; FEATURE:
; OTHER INFORMATION: Amino Acids 116-125-REGION II
; FEATURE:
; OTHER INFORMATION: Amino Acids 126-197-REGION III
; FEATURE:
; OTHER INFORMATION: Amino Acids 198-281-REGION IV
US-09-134-333-14

Query Match      18.7%; Score 424.5; DB 10; Length 281;
Best Local Similarity 35.7%; Pred. No. 8.3e-20;
Matches 131; Conservative 55; Mismatches 56; Indels 115; Gaps 25;

Qy 74 YDVIYKPLAGVYSLKKQIEKNITFTNLNLN-DILNSRLKRRKYFLDVLSDLMQPKHI 132
Db 20 YDVVY-KPLAGMYTKIK-LENHV---NANTNIDMLDSALKKANYFL-VLNSDLN----- 68

Qy 133 SSNEYIIDSFKLLNSQKNTLLSKYIKESVENDIKFAQEGISYVEKVLAKYKDDLES 192
Db 69 PSGEYIHKDPYKLLDLEKK--LGSYKYGASDDT---ANDG--YYKMGLYKH----- 114

Qy 193 IKKVIKEKEKFPSPPTTPSPAKTDOCKESKFLPFLTNIETL---YNNLVNKIDDYI 249
Db 115 ---LVKVEID-----KKGKAKELKYLFLSKQELVKVYTDL----- 149
```

```

Qy 250 INLKAKINDCNVEKDEAHVKITKLSDLKAIDDKID-LFKNPYDFEAIKKLINDDTKKDML 308
Db 150 ---KKINNCOEKKKEV-----KLDYKMDLYKS----- 174

Qy 309 GKLLSTGLVQFPNPTIISKLEGFQDMLNI-----SOHCYKQCPENSGCFRHLDERE 363
Db 175 -KVSSGLLEK-----LMSKLESKL-SLLNVOTQLMSSEHCITDTNVP-NAACYRYLDGTE 227

Qy 364 ECKCLLNKYQEGDKCVENPPTCNENGGCDADATCTEEDSGSSRKKITCECTKPDSPYL 423
Db 228 EWRCLL-FKE--GKVC--PANTC-KDNGGCAPEACKMDN-----IVCKCTREGSEPL 274

Qy 424 FDGIFCS 430
Db 275 FEGVFCs 281

RESULT 19
US-09-804-626-4
; Sequence 4, Application US/09804626
; Patent No. US20020128190A1
; GENERAL INFORMATION:
; APPLICANT: Lobel, Leslie
; APPLICANT: Lustbader, Joyce
; TITLE OF INVENTION: EXPRESSION OF PROPERLY FOLDED AND SOLUBLE EXTRACELLULAR DOMAIN
; FILE REFERENCE: 0575/62259/JPW/SHS
; CURRENT APPLICATION NUMBER: US/09/804,626
; CURRENT FILING DATE: 2001-03-09
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 516
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-09-804-626-4

Query Match      13.5%; Score 306.5; DB 10; Length 516;
Best Local Similarity 27.2%; Pred. No. 5.3e-12;
Matches 124; Conservative 63; Mismatches 156; Indels 113; Gaps 19;

Qy 1 MHHHHHSSGLVPRGSGMKETAATAAKFERQHMDSPDLGTDDDDKAMADIGSIEGRGTMAIS 60
Db 116 MHHHHHSSGLVPRGSGMKETAATAAKFERQHMDSPDLGTDDDDKAMADIGS-EFRVFLCOE 174

Qy 61 VTMDNIIISGFENEYDVIYKPLAGVYSLKK-----QIEKNITFTNLNLNDILNSRLK 113
Db 175 SKVTEIPSDLP--NATELRFVLTKLVIQKAFSGFDLEK---IEISQNDV----- 222

Qy 114 KRKYFLDVLESDL-----MQFKHSSNEYIIEISFKLLNSEQ-----KNTLLKSYKY 160
Db 223 -----LEVIEADVSNLPKLEIRIEKANLLYITPEAFQNLNLOYLIIISNTGKHLPD 277

Qy 161 IKESVENDIKFAQEGISYVEKVLAKYKDDLESIKKVIKEKEKFPSPPTTPSPAKTDE 220
Db 278 VHK-----IHSLQKVLIDQDNI-----NIHTIERNSFVG----- 307

Qy 221 QKESKFLPFLTN-IETLYNNLVNKIDDYIINLKAKINDCNVEKDEAH-----VKI 270
Db 308 LSFESVILWLNKNGIOEIHNAFNGTQLDVNLSDNNNLEELPNDVHFAGSGPVILDIR 367

Qy 271 TKLSDLKAIDDKIDLFKNPYDFAIKKLINDDTKKDMLGKLLSTGLVQFPNPTIISKLE 330
Db 368 TRIHSLPS-----YGLENLKLL-----RARSTVNLKKLPT--LEKLV- 402

Qy 331 GKFDMLNISQHCY-----KQCPENSGCFRHLDERBECKCLLNKYQEGDKCVENPPTC 386
Db 403 ALMEASLTYPSSHCAFAFNWRRQISELHPICNKSILRQEVDMYMTQAGQRSSLAED----- 457

Qy 387 NENGGCDADATCTEEDSGSSRKKITCECT-KPDYS 421
Db 458 NESSYSRGFDWYTFEYDIDLCNEVVDVTCSPKPDFA 493
```





```
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1011
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-896-1011

Query Match      11.5%; Score 262.5; DB 9; Length 355;
Best Local Similarity 49.2%; Pred. No. 2.1e-09;
Matches 60; Conservative 10; Mismatches 19; Indels 33; Gaps 5;

QY 1 MHHHHHSSGLVPRGSGMKETAAAKFERQHMDSPDLGTDDDDKAMADIGSTEGRTWAI- 59
   |||||
Db 116 MHHHHHSSGLVPRGSGMKETAAAKFERQHMDSPDLGTDDDDKAMDRL--VOREGTRAVY 173
   |||||
QY 60 -----SVTMDNLSGFENEYDVIYLPK--LAGVYRSLKKQTEKN 96
   |||||
Db 174 LASVAAPFPAAGATCLSHSVAVVTASAALTGF--TFSAQLILPYTLASLYHR-----EKQ 226
   |||||
QY 97 IF 98
;
Db 227 VF 228

RESULT 23
US-09-681-938-1
; Sequence 1, Application US/09681938
; Publication No. US20030003584A1
; GENERAL INFORMATION:
; APPLICANT: Adams, Darius
; TITLE OF INVENTION: Liposomal Vector Binding Protein for Hepatocyte DNA Delivery
; FILE REFERENCE: 29147
; CURRENT APPLICATION NUMBER: US/09/681,938
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: 09/681,938
; PRIOR FILING DATE: 2001-06-28
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 201
; TYPE: PRT
; ORGANISM: Custom Liposomal Vector Binding Protein for Hepatocyte DNA Delivery
; FEATURE:
; NAME/KEY: BINDING
; LOCATION: (1)..(201)
US-09-681-938-1

Query Match      11.4%; Score 258.5; DB 9; Length 201;
Best Local Similarity 31.6%; Pred. No. 1.9e-09;
Matches 83; Conservative 27; Mismatches 70; Indels 83; Gaps 9;

QY 1 MHHHHHSSGLVPRGSGMKETAAAKFERQHMDSPDLGTDDDDKAMADIGSTEGRTWAI 60
   |||||
Db 1 MHHHHHSSGLVPRGSGMKETAAAKFERQHMDSPDLGTDDDDKAMA----- 46

QY 61 VTMDNLSGFENEYDVIYLPKLAGVYRSLKKQIEKNFTFNLDNLSRLKKRYFLD 120
   |||||
Db 47 -----SPKRSKPRSKPRSGRIPTY---LSE-----D 71

QY 121 VLESDLMQFKHISNEYIIDS--FKLLNSEQKN-----TLKSYKIKESVENDIK 170
   |||||
Db 72 ELKAAEAFKRNHTFEVGHDSGFSITVSGQSPDPQVGIITWMDLKNRGKM--DVS 129

QY 171 FAQEGI-----SYEKKVLAKYKDDLESIKKVIKEKEKFPSPPTTPPSPAK-TDEQKES 225
   |||||
Db 130 GVQAPVGAITTIEDPVLAK-----KVPETFPKPGESRHTSDHMSIY 172

QY 226 KFLPFLNLTNIETLNNLYNKIDDY 248
   |||||
Db 173 KPMGRSHFLCTFTFNNNKKEY 195

RESULT 24
```

```
US-10-042-945-28
; Sequence 28, Application US/10042945
; Publication No. US20030045468A1
; GENERAL INFORMATION:
; APPLICANT: Fling, Steven P.
; APPLICANT: Foy, Teresa M.
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Wang, Aijun
; APPLICANT: Johnson, Jeffrey C.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Sutherland, R. Alec.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY.
; TITLE OF INVENTION: DIAGNOSIS AND MONITORING OF BREAST CANCER
; FILE REFERENCE: 210121.479C3
; CURRENT APPLICATION NUMBER: US/10/042.945
; CURRENT FILING DATE: 2002-01-08
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 55
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-042-945-28

Query Match      7.3%; Score 167; DB 9; Length 55;
Best Local Similarity 100.0%; Pred. No. 0.00027;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 GSGMKETAAAKFERQHMDSPDLGTDDDDKAMA 46
   |||||
Db 1 GSGMKETAAAKFERQHMDSPDLGTDDDDKAMA 32

RESULT 25
US-09-757-417-28
; Sequence 28, Application US/09757417
; Patent No. US20020082216A1
; GENERAL INFORMATION:
; APPLICANT: Fanger, Gary R.
; APPLICANT: Foy, Theresa M.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: THERAPY, DIAGNOSIS AND MONITORING OF BREAST CANCER
; FILE REFERENCE: 210121.479C1
; CURRENT APPLICATION NUMBER: US/09/757.417
; CURRENT FILING DATE: 2001-01-08
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 55
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-757-417-28

Query Match      7.3%; Score 167; DB 10; Length 55;
Best Local Similarity 100.0%; Pred. No. 0.00027;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 GSGMKETAAAKFERQHMDSPDLGTDDDDKAMA 46
   |||||
Db 1 GSGMKETAAAKFERQHMDSPDLGTDDDDKAMA 32

RESULT 26
US-09-842-930A-25
; Sequence 25, Application US/09842930A
; Publication No. US20020197681A1
; GENERAL INFORMATION:
; APPLICANT: Weigel, Paul
; TITLE OF INVENTION: Identification of Hyaluronan Receptor for Endocytosis
; FILE REFERENCE: 5820.603
; CURRENT APPLICATION NUMBER: US/09/842,930A
```

; CURRENT FILING DATE: 2001-04-22  
; PRIOR APPLICATION NUMBER: 60/245,320  
; PRIOR FILING DATE: 2000-11-02  
; PRIOR APPLICATION NUMBER: 60/199,538  
; PRIOR FILING DATE: 2000-04-25  
; NUMBER OF SEQ ID NOS: 56  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 25  
; LENGTH: 1394  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-842-930A-25

Query Match 6.4%; Score 145.5; DB 9; Length 1394;  
Best Local Similarity 19.7%; Pred. No. 0.29;  
Matches 88; Conservative 77; Mismatches 159; Indels 123; Gaps 22;

QY 50 STEGRCTMAISVTMDNLSGFENEVDYVILKPLAGVYRSLLKQIEKNITFFNLNL-----104  
DB 8 AIEADAYTVFAPNNAIENYIREKKVL-----SLEDVILYHVLEEKLL 53  
QY 105 -NDILNSRLKKR-----KYFLD-VLSEDLN-----QFKHISSEYIIEDSEKLLNSEOK 151  
DB 54 KNDLNGMHRETMLGFSYFLSFELHNDQLYVNEAPINYTNVATDKGVIQ--YNLANAIEA 111  
QY 152 NTLKSYKYIKESVENDIKFAQEGISYVEKVLAKYKDDLESIKKVIKEEKXPPSPPTT 211  
DB 112 ADAYTVFAPNNAIENYIR-ERKVLSEEDVL-RYHVLE--EKLKNDLH-----158  
QY 212 PPSPAKTDOEKESKFLPFLTNIETLYNNLVNKIDIDLINLAKINDCNVDEAHVKIT 271  
DB 159 --NGMHRETMLGFSYFLSFELHNDQLYV-----EAPINYTNVATDKGVIQYN 204  
QY 272 KLSLDKAID-----DKI-----DLFKNPYDFAIKKLINDDT-----303  
DB 205 LANATEAADAYTVFAPNNAIENYIREKKVLSLEEDVLR--YHVLEEKLLNDLNGMH 262  
QY 304 KKDMLG-----KLLSTGLVQNPFPNTIISKLI-----EGKFQDMLNISQHCQVK 346  
DB 263 RETMLGFSYFLSFELHNDQLYVNEAPINYTNVATDKGVCAAGFQGGTICTTAINACEISN 322  
QY 347 KCPENSGCFRHLDEREECKLLNKKQEGDKVE-NPNTCTENNGGCCDADATCTEEDSG 405  
DB 323 GCSAKADCKRTTPGRVRCTKAGYTGOGIVCLEINP-----CLENHGGCDKNAECTQ-----375  
QY 406 SSRKKITCECTKPDSPYPLF--DGIFCS 430  
DB 376 TGPNOACNC-----LPATYTGDKVCT 397

RESULT 27  
US-09-815-242-5251  
; Sequence 5251, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows version 4.0  
; SEQ ID NO 5251  
; LENGTH: 996  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-5251

Query Match 6.3%; Score 144; DB 10; Length 996;  
Best Local Similarity 20.4%; Pred. No. 0.24;  
Matches 93; Conservative 70; Mismatches 133; Indels 161; Gaps 21;  
QY 25 KFEROHMDSPDLGTDGDD-----DKAMADIGSIEGRGTMAISVTMDNLSGFE 71  
DB 387 KQOOSTLNIERTLDEQGLDELNNLIKLEKVEDSIGNESDYKILIE--LNNAITNIN 444  
QY 72 NEYDVI-----YKPLAGVYRSLLKQI--EKNIF-----TFNLNLD-----106  
DB 445 NEINVIKENAKADELDKLLGSKQELENOINEEKTILKNLEIKLDYDKSKLDLNDKESF 504  
QY 107 -----LNSRLKKRYFLDY--LESOLMQFKHISSEYIIEDSEKLLNSE 149  
DB 505 ISEIKSAVKIGDQPCGNEIQDLGHIDFDSIAKQNETKEATEANTHTMESIAVHNS 564  
QY 150 QKNTLLSKYIKESVEN-DIKFAQEGISYVEKVLAK-----YKDDLESIKKVIK 198  
DB 565 -----IKFVNEKISNINIK--TQSDLSL--EVLNKRLENNALNNQDLNKFIEQMK 613  
QY 199 EEKEKPPSPPTTPPSPAKTDEQKESKFLPFLTNIETL--YNNLV-----242  
DB 614 EEKDNLTQLHINKQLKLNKNESELKICRNL--ITEFETLSKYNNITNFEVDYKYYIQDYN 671  
QY 243 -----NKIDDLVILNKAK-----INDCNVDEAHVKITKL 273  
DB 672 QHQSINQIEDKLIQLSQRKLTBQNNLHYNQLETYNNNDLENEQSIEMEMSRNLTD 731  
QY 274 SLDKAI-----DDKIDLFKNPY-DPE-----ATKKLINDDTKKD---ML 308  
DB 732 NDINEIARWGQOELEQRDRDYKKRYHEFEMEIALESITKDKLELDSKUKADDYELKK 791  
QY 309 GKLLSTGLVQNPFPNTIISKLEKGFQDMLNISQHCQV 345  
DB 792 GKM-----NTLIDEVSAVHYOCNNIKKTQSI 818

RESULT 28  
US-09-815-242-12141  
; Sequence 12141, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12141  
; LENGTH: 1009  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-12141

Query Match 6.3%; Score 144; DB 10; Length 1009;  
Best Local Similarity 20.4%; Pred. No. 0.25;  
Matches 93; Conservative 70; Mismatches 133; Indels 161; Gaps 21;

Qy 25 KFERHMDSPDLGTDGDD-----DKAMADIGSIEGRGTMAISVTMDNLISGFE 71  
Db 387 KYQSYLNIERTDEQGLBELNLKLGLEKVEDSIGNSDYEKIIE--LNNAITNIN 444  
Qy 72 NEYDVI-----YLKPLAGYRSLKKQI--EKNIF-----TFNLNLND----- 106  
Db 445 NEINVIKENAKDELKLLGSKOLENQINEKTLANKLEIKLDYDKSKLDLNDKESF 504  
Qy 107 -----ILNRLKRRKYFDV--LESDLMOFKHISNEYIIEDSFKLLNSE 149  
Db 505 ISEIKSAVKIGDQPCIGNEIQDLGHDFDSIAKQRONEIKIEIANIHTMESNIAVHNS 564  
Qy 150 QKNTLLSKYKIKESVEN-DIKFAQEGISYKVLAK-----YKDDLESIKKVIK 198  
Db 565 -----IKFVNEKISINIK--TQSDLSL--EVLNKRLEENALNNQRLNKFIEQMK 613  
Qy 199 BEKEKFPSPPTPPSPAKTDEQKESKFLPLTNIEFL--YNNLV----- 242  
Db 614 BEKONLTQIHNKQLRLKNESELKICRNL--ITEFETLSKYNNITNEVDYKKYIQDVN 671  
Qy 243 -----NKIDDYLLNLKAK-----INDCNVEKDEAHVKITKL 273  
Db 672 QHQHSNQIEDKLIQLSQRKLEQNNLNHYENQLETYNNDLENEQSIEMEMSRNLTTDD 731  
Qy 274 SLDKAI-----DDKIDLFKNPY--DFE-----AIKKLINDDTKDD-----ML 308  
Db 732 NDINEIINAWRGEQELEQKRDYTKRRYHEFEMETARLESITKDELSDSKLDDYELKK 791  
Qy 309 GKLLSTGLVQFPNTIISKLEGFQDMNLNISOHQCV 345  
Db 792 GKM-----NTLIDEYSVAVHYQCCNNLKKTKQSI 818

RESULT 29  
US-09-888-615-65  
; Sequence 65, Application US/09888615  
; Patent No. US20020064856A1  
; GENERAL INFORMATION:  
; APPLICANT: PLOWMAN, GREGORY  
; APPLICANT: WHYTE, DAVID  
; APPLICANT: CAENEPEEL, SEAN  
; APPLICANT: CHARYDCZAK, GLEN  
; APPLICANT: MANNING, GERARD  
; APPLICANT: SUDARSANAM, SUCHA  
; TITLE OF INVENTION: NOVEL PROTEASES  
; FILE REFERENCE: 038602/1214  
; CURRENT APPLICATION NUMBER: US/09/888,615  
; CURRENT FILING DATE: 2001-06-26

; PRIOR APPLICATION NUMBER: 60/214,047  
; PRIOR FILING DATE: 2000-06-26  
; NUMBER OF SEQ ID NOS: 150  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 65  
; LENGTH: 980  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-888-615-65

Query Match 6.2%; Score 140.5; DB 10; Length 980;  
Best Local Similarity 20.9%; Pred. No. 0.4;  
Matches 77; Conservative 64; Mismatches 154; Indels 73; Gaps 15;

Qy 67 LSGFENEYDVILKPLAGYRSLKKQIEKNITFNILNLNDILNLSRLKRRKYFDLVLESDL 126  
Db 341 LOGFSNLGNTCYM-----NAILQSLSFLOSFANDLLKQIPWKKIPLNAL---I 386  
Qy 127 MOFKHISNEYIIEDSFKLLNSEQKNTLLSKYKIKESVENDIKFPAQEGISYY----- 179  
Db 387 RRFHALLVKK-----DICNSETKKDLK-----KVKNNAISATAERFSYGMQDAHE 432  
Qy 180 --EKVLAKYKDDLESIKKVIKEKEKFFSPPTTPPSAKTDEQKESKFLPLTNIE-T 236  
Db 433 FLSQCLDQKEDMEKLNKTWKTEPVSGEENSPDISATRAYT-----CPVITNLFE 483  
Qy 237 LYNLVNKIDYILNLKAKINDCNVEKDEAHVKITKLSDLKAIKIDLFKNPYDFE-AI 295  
Db 484 VOHSIICKACGEIIPKREQFNLDLIDLPKPKLPP-----RSIQDSLDFPRAEELEYSC 539  
Qy 296 KK-----LINDDTYKMDLGLLSTGLVQNPNTIISKLEGFQDMNLNISOHQCVKQC 349  
Db 540 EKCGGKCALVRH--KFNRLPRVILHLKRYSFNVALS--LNNKIGQVVIIPRYLTSSHC 595  
Qy 350 PENS-----GCFRHLDERECKLLNFKQEGDKCVENPN-PTCNENNGGCDADATCTEE 402  
Db 596 TENTKPPFTLGSAAHMAHMRPLKA-----SQMVNSCITSPSPSKFTFKSKSSLLCLDS 651  
Qy 403 DSGSSRKK 410  
Db 652 DSEDELR 659

RESULT 30  
US-10-267-311-21  
; Sequence 21, Application US/10267311  
; Publication No. US20030050469A1  
; GENERAL INFORMATION:  
; APPLICANT: Siegel, Marvin  
; APPLICANT: Chu, N. Randall  
; APPLICANT: Mizzen, Lee A.  
; TITLE OF INVENTION: INDUCTION OF A THI-LIKE RESPONSE IN VITRO  
; FILE REFERENCE: 12071/002001  
; CURRENT APPLICATION NUMBER: US/10/267,311  
; CURRENT FILING DATE: 2002-10-09  
; PRIOR APPLICATION NUMBER: US/09/613,303  
; PRIOR FILING DATE: 2000-07-10  
; PRIOR APPLICATION NUMBER: US 60/143,757  
; PRIOR FILING DATE: 1999-07-08  
; NUMBER OF SEQ ID NOS: 55  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 21  
; LENGTH: 948  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: fusion sequence  
US-10-267-311-21

Query Match 5.8%; Score 131.5; DB 9; Length 948;  
Best Local Similarity 20.0%; Pred. No. 1.4;  
Matches 82; Conservative 62; Mismatches 123; Indels 143; Gaps 16;



```
; Sequence 2239, Application US/10146502
; Publication No. US20030069180A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secret, Heather
; APPLICANT: Wang, Aijun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.527C2
; CURRENT APPLICATION NUMBER: US/10/146.502
; CURRENT FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2241
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2239
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-146-502-2239

Query Match      5.6%; Score 128; DB 9; Length 391;
Best Local Similarity 18.8%; Pred. No. 0.83;
Matches 82; Conservative 64; Mismatches 171; Indels 120; Gaps 16;

Qy  2 HHHHHHSGGLVPRGSMKETAATAAFERQHMDSPLDGTDDDKAMADIGSIEGRGTMAISV 61
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  5 HHHHHHSGGLVPRGSHM-----DARRVPQKDLRVKLNKLFYVKLI-----SMETSS 52

Qy  62 TMDNLSCFENEYDVIYKPLAGVYRSLSKQIEKNIFTNLDNLNLSRLKRRKRYFDV 121
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  53 SSDSCSDFASD-----NFANRLQ-----SV 74

Qy  122 LESDLMQFKHSSNEYIEDSFKLLNSP-QKNTLLKSYKIKESV--ENDIKFAQEGISY 178
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  75 REGCTRQCRHSGLRVAMKFPARGATNKKAESQPSNSVTSNDSSEDESGMNF 134

Qy  179 YEKVLAKYKDDLSIKVKEKEKFPSSPTTPPSPAKTDEQK----- 223
    || : : : : : || : : : : : || : : : : : || : : : : : || : : : :
Db  135 LEKRALNIKONKAKLKM-SELESFPGSFRGRLPLGSDSQSRPRRRTPPGVASRRNP 193

Qy  224 ESKFLP-----FLTNIETLYNNLVNKIDYLLNLKAKINDCNVEKDEAHVKITKLSDL 276
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  194 ERRARLTRSRRLGSLDALPMEEEDYMLVRKRTVDGYMNEDD-----LPRS 246

Qy  277 KAIDDKIDLFKNPYDFAIKLLINDT7KKOMLGLSTGLVQNPNTIISKLEGKFQDM 336
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  247 RRSRSVTL---PHIRPVEITEELE-----NVCSNSREKIYNS 285

Qy  337 LNISSOHCVKQCPENSGCFRHLDEREC-----KCLLN-YKQGGDKCVENPNPTC 386
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  286 LGSTCHOCROKTIOTKNTC-----RNPDCMGVGFQCGPCPLRNRYGVEVRDALLDPNWHC 340

Qy  387 NENGGCDADATCEED 403
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  341 PPCRGICNC-SFCRORD 356

RESULT 34
US-09-764-864-1160
; Sequence 1160, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT23
; CURRENT APPLICATION NUMBER: US/09/764.864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1160
; LENGTH: 580

; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-764-864-1160

Query Match      5.6%; Score 127.5; DB 9; Length 690;
Best Local Similarity 23.8%; Pred. No. 1.8;
Matches 81; Conservative 47; Mismatches 145; Indels 67; Gaps 13;

Qy  18 MKETAAAKFERQHMDSP-----DLGTDD--DDKAMADIGSIEGRGTMAISVTMDN 65
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  245 LKEAVENKATSEODKPKRAKRGVSGELATPDKKENDAKSSDSVGEETLPSPSLNAN 304

Qy  66 ILSGFENEYDV- IYKPLAGVYRSLSKQIEKNIFTNLDNLNLSRLKRRKRYFDVLES 124
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  305 -ESQTEHRKDVDEYIKKMLS-----EIQLDRRKHTQNVNLKLSAIRTLYLSVLKE 358

Qy  125 D-----LMQPKHISSNEVIIEISFKLL-----NSQKNTLL 155
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  359 NSKKEELTSKTKAELTAAPQFKKDTLKPEKKAEEAKKVEEAKKAKDQKEDRRNPT 418

Qy  156 KSYKIKESVENDIKFAQEGISYKVLAKYKDDLSIKVKEKEKFPSS-SPPTTTPPS 214
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  419 NTKTLELEAESDVK-----VKEAELELVKEEANSNEEKIKQAKEKVESKKAERLE 473

Qy  215 PAKTDEQKESKFLPFLTNIETLYNNLVNKIDYLLNLKAKINDCNVEKDEAHVKITKL- 273
```



Db	268	-ESOTEHRKVDVEYIKKMLS-----ETQLDRRHQTQNVNINIKLSAITKTYLYELSVLKE	321
Qy	125	D-----LMQPKHISSNEYIIEDSFKLL---NSEQKNLL	155
Db	322	NSKKEELTSKTAELTAAEQFKDILKPEKKVAEAEKKVVEAKKKAKDOKEEDRRNYPT	381
Qy	156	KSYKYYK-ESVENDIKPAQGGISYYEKVLAKYKDDLESIKKVKYKEEKPPS-SPPTTTP	213
Db	382	NTYKLTLEIAESDVK---VKRAEELVELYKEEANESRNEEKIKQAKKEKVESKAAEATRL	436
Qy	214	SPAKTDQKKESKFLPFLTNIETLYNNLVNKIDYLLNLKAKINLCNDCHNVEDEAHVKITKL	273
Db	437	EKIKDRKKAEEAEAKRAESEKKAFAKOKVDAEEAFAEAKTAELYE-----VQRL	489
Qy	274	-SDUKAID--DKIDLFKNPYDFEAIKKLINDDTKKMLGKL	311
Db	490	EKELKEIDESDSEYLYKEGLRAPLQSKL---DTKKAKLSKL	527

RESULT 40  
 US-09-298-523B-1  
 ; Sequence 1, Application US/09298523B  
 ; Publication No. US20030059438A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: BRILES et al.  
 ; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS  
 ; TITLE OF INVENTION: AND STRAINS THEREOF AND USES THEREFOR  
 ; FILE REFERENCE: 454312-3140  
 ; CURRENT APPLICATION NUMBER: US/09/298,523B  
 ; CURRENT FILING DATE: 1999-04-23  
 ; NUMBER OF SEQ ID NOS: 78  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 1  
 ; LENGTH: 691  
 ; TYPE: PRT  
 ; ORGANISM: Streptococcus pneumoniae  
 US-09-298-523B-1

Query Match	5.4%	Score 123;	DB 9;	Length 691;
Best Local Similarity	24.0%;	Pred. No. 3.4;		
Matches 82;	Conservative	47;	Mismatches 144;	Indels; 68;
			Gaps	14;

```

245 LKAEAVKNVTSODKPKRRKRGVSGELATPOKKENDAKSSDSSVGEETLPSPSLMAN 304
QY 66 ILSGFENEYDV-IYLPKPLAGVYSLKKQIEKNFIFTNLNLINDLSNRKKRYFLDWLLES 124
Db 305 -ESOTEHRKDVDEYIKKMLS-----ETQLDRRHQHTQVNLNLKLSAIKTKYLYSELSVLKE 358
QY 125 D-----LMQPKHSSNEVYIIEISFKLL-----NSEOKNTLL 155
Db 359 NSKKEELTSKTKAELTAAPQFPKDTLLKPEKVAEAEKKVVEEAKKKAKDOEEDRRNYPT 418
QY 156 KSYKYIK-ESVENDIKFAQGGISYVEKVLAKYKDDLESIKKVIKEKEKEPPS-SPPTTPP 213
Db 419 NTYKTLLEIAESDVK-----VKEAELELVKEEANESRNEEKIKQAKEKVESKKAETRL 473
QY 214 SPARTDQKESKESFLPPLTNIETLYNNLVNKRIDYLLNLKAKINDCNVDEAHVKTTL 273
Db 474 EKIKTRKKEEAEKAKRKAEESEKKAEEAKQKVDAAEEVALEAKIALEYE-----VQRL 536

```

```

QY 274 -SOLKATID--DKIDLFKNPYDFAIRKKLIINDDTKKDMLGKL 311
      :||| || | | | | | | | | | | | | | | | | | | | |
Db 527 EKELKEIDESESDYLKEGLRAPLQSKL---DTKKAKLSKL 564

```

Search completed: May 12, 2003, 10:25:59  
Job time : 58 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 10:15:38 ; Search time 46 seconds  
(without alignments)  
900.738 Million cell updates/sec

Title: US-10-057-531A-2

Perfect score: 2275

Sequence: 1 MHHHHHHSSGLVPRGSMKE.....TCECTKPDSPYPLFDGIFCSS 431

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_73:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1946.5	85.6	1701	2 A54498	major merozoite su
2	1935.5	85.1	1726	1 S42QGM	major merozoite su
3	1925.5	84.6	1701	2 A26868	major merozoite su
4	1917	84.3	651	2 S47282	merozoite surface
5	1916.5	84.2	1726	2 A45948	major merozoite su
6	1175	51.6	400	2 A45545	major merozoite su
7	1175	51.6	400	2 S05603	major merozoite su
8	1051.5	46.2	1631	1 S42QK1	major merozoite su
9	1007.5	44.3	1640	2 A24594	probable major sur
10	816	35.9	1726	2 A39401	merozoite surface
11	793.5	34.9	1751	2 A45604	major blood-stage
12	556.5	24.5	680	2 A28121	major merozoite su
13	556.5	24.5	1772	2 A45532	major merozoite su
14	552	24.3	1785	2 A45546	major merozoite su
15	180.5	7.9	980	2 E71606	hypothetical prote
16	167.5	7.4	2269	2 T28677	rhoptry protein -
17	166	7.3	2166	2 G70163	hypothetical prote
18	160.5	7.1	1169	2 A64505	pl15 homolog - Met
19	156.5	6.9	1939	2 T18372	repeat organellar
20	154.5	6.8	1156	2 B70356	chromosome assembl
21	152	6.7	1191	2 B97116	chromosome segrega
22	152	6.7	2401	2 T28676	rhoptry protein -
23	149.5	6.6	652	2 B59102	hypothetical prote
24	147	6.5	909	2 C97325	hypothetical prote
25	146.5	6.4	622	2 S51972	SPC72 protein - ye
26	146	6.4	1619	2 T18499	hypothetical prote
27	146	6.4	3724	2 T18427	hypothetical prote
28	145	6.4	1130	2 T34081	hypothetical prote
29	144	6.3	671	2 H64502	hypothetical prote

30	144	6.3	3394	2 T18501	hypothetical prote
31	143.5	6.3	1086	2 S16752	major merozoite su
32	143	6.3	540	2 S38085	hypothetical prote
33	142.5	6.3	864	2 B90395	purine Nriase [imp
34	142.5	6.3	963	2 C90535	conserved hypotet
35	142	6.2	1005	2 A64465	hypothetical prote
36	142	6.2	1187	2 T18355	hypothetical prote
37	142	6.2	1679	2 S48385	hypothetical prote
38	142	6.2	2819	2 A90551	conserved hypotet
39	140.5	6.2	819	2 E70105	pl15 protein homol
40	140	6.2	1009	2 C89910	hypothetical prote
41	140	6.2	1127	2 T28317	ORF MSV156 hypotet
42	140	6.2	1712	2 C71618	hypothetical prote
43	139.5	6.1	1233	2 T40059	chromosome segrega
44	139.5	6.1	1979	2 C71622	hypothetical prote
45	139	6.1	1188	2 A71621	protein with 5'-3'

ALIGNMENTS

RESULT 1

A54498

major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum

C:Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 09-Jun-2000

C:Accession: A54498

R:Peterson, M.G.; Coppel, R.L.; McIntyre, P.; Langford, C.J.; Woodrow, G.; Brown, G.V

Mol. Biochem. Parasitol. 27, 291-302, 1988

A:Title: Variation in the precursor to the major merozoite surface antigens of Plasmo

A:Reference number: A54498; MUID:88142999; PMID:2449612

A:Accession: A54498

A>Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-1701 <PEP>

A:Cross-references: GB:M19143; NID:q160412; PIDN:AAA29653.1; PID:q160413

C:Superfamily: major merozoite surface antigen

C:Keywords: surface antigen

Query Match 85.6%; Score 1946.5; DB 2; Length 1701;

Best Local Similarity 94.9%; Pred. No. 4.9e-92;

Matches 374; Conservative 7; Mismatches 8; Indels 5; Gaps 1;

QY	38	TDDDDKAMADIGSIEGRGTMAISVTMDNLTLSGFENEDVYTLKPLAGVYSLKKQIEKNI	97
DB	1293	SEDNDEYLDQVVTGE-----AISVTMDNLTLSGFENEDVYTLKPLAGVYSLKKQIEKNI	1347
QY	98	FTFNLMNDILNSRLKRRKYFLDVLSDLMQFKHISSEYIIEDSPKLLNSEQKNTLLKS	157
DB	1348	ITFNLMNDILNSRLKRRKYFLDVLSDLMQFKHISSEYIIEDSPKLLNSEQKNTLLKS	1407
QY	158	YKIKESVENDIKFAQEGISYEKVLAKYKDDLESIKKVIKEKEKFPSPPTPPSPAK	217
DB	1408	YKIKESVENDIKFAQEGISYEKVLAKYKDDLESIKKVIKEKEKFPSPPTPPSPAK	1467
QY	218	TDEQKESFLPPLTNIETLYNNLVNKIDDLINLAKINDCNVDEKDEAHVKTIKLSDLK	277
DB	1468	TDEQKESFLPPLTNIETLYNNLVNKIDDLINLAKINDCNVDEKDEAHVKTIKLSDLK	1527
QY	278	AIDDKIDLKPNVDFPAIKKLINDDTKKMDLGLLSTGLVQNFNPNTIISKLEGGKFDML	337
DB	1528	AIDDKIDLKPNVDFPAIKKLINDDTKKMDLGLLSTGLVQNFNPNTIISKLEGGKFDML	1587
QY	338	NTISOHCVKQCCPNSGCFRHLDERECCLLNYKQEGDKCVENPNTCTNENNGGCADA	397
DB	1588	NTISOHCVKQCCPNSGCFRHLDERECCLLNYKQEGDKCVENPNTCTNENNGGCADA	1647
QY	398	TCTEEDSGSSRKKITCECTKPDSPYPLFDGIFCSS	431
DB	1648	TCTEEDSGSSRKKITCECTKPDSPYPLFDGIFCSS	1681

RESULT 2

## SA2QGM

major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (strain RO-71)  
N:Alternate names: 195K glycoprotein  
C:Species: Plasmodium falciparum  
C:Date: 30-Sep-1987 #sequence\_revision 31-Mar-1991 #text\_change 09-Jun-2000  
C:Accession: A2386; S06361  
R:Weber, J.L.; Leininger, W.M.; Lyon, J.A.  
Nucleic Acids Res. 14, 3311-3323, 1986  
A:Title: Variation in the gene encoding a major merozoite surface antigen of the human merozoite  
A:Reference number: A23386; MUID:86205236; PMID:3517809  
A:Accession: A23386  
A:Molecule type: DNA  
A:Residues: 1-1104 <WEB>  
A:Cross-references: EMBL:X03831  
R:Weber, J.L.; Sim, B.K.L.; Lyon, J.A.; Wolff, R.  
Nucleic Acids Res. 16, 1206, 1988  
A:Title: Merozoite surface protein sequence from the Camp strain of the human malaria parasite  
A:Reference number: S06361; MUID:88143999; PMID:3278296  
A:Accession: S06361  
A:Molecule type: DNA  
A:Residues: 1104-1726 <WEB2>  
A:Cross-references: EMBL:X03831  
C:Comment: The merozoite stages of different strains have strain-specific surface antigens  
C:Comment: P. falciparum has three stages: sporozoite, merozoite, and gametocyte. The merozoite stage has three surface antigens  
C:Superfamily: major merozoite surface antigen  
C:Keywords: glycoprotein; merozoite; surface antigen; tandem repeat  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-1726/Product: major merozoite surface antigen #status predicted <MAT>  
F:67-87, 91-96, 100-105, 109-120/Region: 3-residue repeats (S-G-T)  
F:757-765/Region: 3-residue repeats (T-E-E)  
F:133, 272, 501, 567, 638, 827, 839, 924, 944, 990, 1016, 1114, 1221, 1613, 1658/Binding site: carboxy

Query Match 85.1%; Score 1935.5; DB 1; Length 1726;  
Best Local Similarity 94.4%; Pred. No. 1.8e-91;  
Matches 372; Conservative 8; Mismatches 9; Indels 5; Gaps 1;

Qy 38 TDDDDKAMADIGSTEGRTMAISVTMDNILSGFENEYDVIVLKLPLAGYVRSLLKQIEKNI 97  
Db 1318 SEDNDEYLDQVVTGE-----AISVTMDNILSGFENEYDVIVLKLPLAGYVRSLLKQIEKNI 1372  
Qy 98 FTFNLNLDILNSRLKRRKYPFLDVLSDLMQFKHISSEYIIEDSFKLLNSEQNTLLKS 157  
Db 1373 FTFNLNLDILNSRLKRRKYPFLDVLSDLMQFKHISSEYIIEDSFKLLNSEQNTLLKS 1432  
Qy 158 YKIKESVENDIKFAQEGISYEVYKAKYKDDLESIKKVIKEEKEKFPSPPTPPSPAK 217  
Db 1433 YKIKESVENDIKFAQEGISYEVYKAKYKDDLESIKKVIKEEKEKFPSPPTPPSPAK 1492  
Qy 218 TDEQKESKFLPFLTNIETLNNLVNKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLK 277  
Db 1493 TDEQKESKFLPFLTNIETLNNLVNKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLK 1552  
Qy 278 AIDDKIDLFPKNYDFEATKLLINDTKMDLGLKLLSTGLGVNFPNTIISKLIIEGKFQDML 337  
Db 1553 AIDDKIDLFPKNYDFEATKLLINDTKMDLGLKLLSTGLGVNFPNTIISKLIIEGKFQDML 1612  
Qy 338 NISOHQCVKQCPNSGCFRHLDEREECKLLNYKQEGDKCVENPNPTCNENNGCCDADA 397  
Db 1613 NISOHQCVKQCPNSGCFRHLDEREECKLLNYKQEGDKCVENPNPTCNENNGCCDADA 1672  
Qy 398 TCTEEDSGSSRRKTKTCECTKPDSPYPLDFGIFCSS 431  
Db 1673 KCTEEDSGSGKTKTCECTKPDSPYPLDFGIFCSS 1706

## RESULT 3

A26868  
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (strain RO-71)  
C:Species: Plasmodium falciparum  
C:Date: 19-Nov-1988 #sequence\_revision 19-Nov-1988 #text\_change 09-Jun-2000  
C:Accession: A26868  
R:Tanabe, K.; Mackay, M.; Goman, M.; Scalfe, J.G.  
J. Mol. Biol. 195, 273-287, 1987

A:Title: Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmodium falciparum  
A:Reference number: A26868; MUID:88011243; PMID:3079521  
A:Accession: A26868  
A:Molecule type: DNA  
A:Residues: 1-1701 <TAN>  
C:Superfamily: major merozoite surface antigen  
C:Keywords: signal sequence  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-1701/Product: major merozoite surface antigen #status predicted <MAT>

Query Match 84.6%; Score 1925.5; DB 2; Length 1701;  
Best Local Similarity 94.2%; Pred. No. 5.8e-91;  
Matches 371; Conservative 7; Mismatches 11; Indels 5; Gaps 1;

Qy 38 TDDDDKAMADIGSTEGRTMAISVTMDNILSGFENEYDVIVLKLPLAGYVRSLLKQIEKNI 97  
Db 1293 SEDNDEYLDQVVTGE-----AISVTMDNILSGFENEYDVIVLKLPLAGYVRSLLKQIEKNI 1347  
Qy 98 FTFNLNLDILNSRLKRRKYPFLDVLSDLMQFKHISSEYIIEDSFKLLNSEQNTLLKS 157  
Db 1348 FTFNLNLDILNSRLKRRKYPFLDVLSDLMQFKHISSEYIIEDSFKLLNSEQNTLLKS 1407  
Qy 158 YKIKESVENDIKFAQEGISYEVYKAKYKDDLESIKKVIKEEKEKFPSPPTPPSPAK 217  
Db 1408 YKIKESVENDIKFAQEGISYEVYKAKYKDDLESIKKVIKEEKEKFPSPPTPPSPAK 1467  
Qy 218 TDEQKESKFLPFLTNIETLNNLVNKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLK 277  
Db 1468 TDEQKESKFLPFLTNIETLNNLVNKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLK 1527  
Qy 278 AIDDKIDLFPKNYDFEATKLLINDTKMDLGLKLLSTGLGVNFPNTIISKLIIEGKFQDML 337  
Db 1528 AIDDKIDLFPKNYDFEATKLLINDTKMDLGLKLLSTGLGVNFPNTIISKLIIEGKFQDML 1587  
Qy 338 NISOHQCVKQCPNSGCFRHLDEREECKLLNYKQEGDKCVENPNPTCNENNGCCDADA 397  
Db 1588 NISOHQCVKQCPNSGCFRHLDEREECKLLNYKQEGDKCVENPNPTCNENNGCCDADA 1647  
Qy 398 TCTEEDSGSSRRKTKTCECTKPDSPYPLDFGIFCSS 431  
Db 1648 TCTEEDSGSSRRKTKTCECTKPDSPYPLDFGIFCSS 1681

## RESULT 4

S47282  
merozoite surface antigen 1 - malaria parasite (Plasmodium falciparum) (strain RO-71)  
C:Species: Plasmodium falciparum  
A:Variety: strain RO-71  
C:Date: 06-Jan-1995 #sequence\_revision 26-Jul-1996 #text\_change 09-Jun-2000  
C:Accession: S47282  
R:Tolle, R.; Bujard, H.; Cooper, J.A.  
submitted to the EMBL Data Library, July 1994  
A:Description: Plasmodium falciparum: recombination within the C-terminal region of merozoite surface antigen  
A:Reference number: S47282  
A:Accession: S47282  
A:Molecule type: DNA  
A:Residues: 1-651 <TOL>  
A:Cross-references: EMBL:Z35329; NID:g535257; PIDN:CAA84558.1; PID:g535258  
A:Experimental source: strain RO-71  
C:Superfamily: major merozoite surface antigen  
C:Keywords: glycoprotein; merozoite; surface antigen

Query Match 84.3%; Score 1917; DB 2; Length 651;  
Best Local Similarity 93.9%; Pred. No. 5.3e-91;  
Matches 370; Conservative 9; Mismatches 9; Indels 6; Gaps 2;

Qy 38 TDDDDKAMADIGSTEGRTMAISVTMDNILSGFENEYDVIVLKLPLAGYVRSLLKQIEKNI 97  
Db 244 SEDNDEYLDQVVTGE-----AISVTMDNILSGFENEYDVIVLKLPLAGYVRSLLKQIEKNI 298  
Qy 98 FTFNLNLDILNSRLKRRKYPFLDVLSDLMQFKHISSEYIIEDSFKLLNSEQNTLLKS 157  
Db 299 FTFNLNLDILNSRLKRRKYPFLDVLSDLMQFKHISSEYIIEDSFKLLNSEQNTLLKS 358

QY 158 YKIKESVENDIKFAQEGISYEKVLAKYKDDLESIKKVIKEEKEFPSSPTTPSPAK 217  
|||||  
Db 359 YKIKESVENDIKFAQEGISYEKVLAKYKDDLESIKKVIKEEKE-FPSSPTTPSPAK 417  
QY 218 TDEQKESKFLPFLNIETLYNNLVNKIDYILINLAKKINDCNVKEDEAHVKITKLSDLK 277  
Db 418 TDEQKESKFLPFLNIETLYNNLVNKIDYILINLAKKINDCNVKEDEAHVKITKLSDLK 477  
QY 278 ATDDKIDLFKNPYDEAIAKKLNDOTKKDMLGKLLSTGLVQNPNTIISKLEGGFQDML 337  
Db 478 ATDDKIDLFKNPYDEAIAKKLNDOTKKDMLGKLLSTGLVQNPNTIISKLEGGFQDML 537  
QY 338 NISQHCVKKQCPENSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENNGGCDADA 397  
Db 538 NISQHCVKKQCPENSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENNGGCDADA 597  
QY 398 TCTEEDSGSSRKITCECTKPDSPYPLFDGIFCSS 431  
Db 598 KCTEEDSGSGNGKKITCECTKPDSPYPLFDGIFCSS 631

RESULT 5  
A:5948  
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (st  
C:Species: Plasmodium falciparum  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 09-Jun-2000  
C:Accession: A45948  
R:Chang, S.P.; Kramer, K.J.; Yanaga, K.M.; Kato, A.; Case, S.E.; Siddiqui, W.A.  
Exp. Parasitol. 67, 1-11, 1988  
A:Title: Plasmodium falciparum: gene structure and hydropathy profile of the major merozo  
A:Reference number: A45948; MUID:89005525; PMID:3049134  
A:Accession: A45948  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1726 <CHA>  
A:Cross-references: GB:M37213  
C:Superfamily: major merozoite surface antigen  
C:Keywords: surface antigen

Query Match 84.2%; Score 1916.5; DB 2; Length 1726;  
Best Local Similarity 93.7%; Pred. No. 1:7e-90;  
Matches 369; Conservative 8; Mismatches 12; Indels 5; Gaps 1;  
QY 38 TDDDDKAMADIGSIEGRGTMAISVTMDNLSGFENEYDVIYKPLAGVYRSLLKQIEKNI 97  
Db 1318 SEDNDEYLDQVVTGE-----AISVTMDNLSGFENEYDVIYKPLAGVYRSLLKQIEKNI 1372  
QY 98 FTFNLNLDILNSRLKRRKYFDLVLESQKPKHISSEYIETDSFKLLNSQKNTLLKS 157  
Db 1373 FTFNLNLDILNSRLKRRKYFDLVLESQKPKHISSEYIETDSFKLLNSQKNTLLKS 1432  
QY 158 YKIKESVENDIKFAQEGISYEKVLAKYKDDLESIKKVIKEEKEFPSSPTTPSPAK 217  
Db 1433 YKIKESVENDIKFAQEGISYEKVLAKYKDDLESIKKVIKEEKEFPSSPTTPSPVK 1492  
QY 218 TDEQKESKFLPFLNIETLYNNLVNKIDYILINLAKKINDCNVKEDEAHVKITKLSDLK 277  
Db 1493 TDEQKESKFLPFLNIETLYNNLVNKIDYILINLAKKINDCNVKEDEAHVKITKLSDLK 1552  
QY 278 ATDDKIDLFKNPYDEAIAKKLNDOTKKDMLGKLLSTGLVQNPNTIISKLEGGFQDML 337  
Db 1553 ATDDKIDLFKNPYDEAIAKKLNDOTKKDMLGKLLSTGLVQNPNTIISKLEGGFQDML 1612  
QY 338 NISQHCVKKQCPENSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENNGGCDADA 397  
Db 1613 NISQHCVKKQCPENSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENNGGCDADA 1672  
QY 398 TCTEEDSGSSRKITCECTKPDSPYPLFDGIFCSS 431  
Db 1673 KCTEEDSGSGNGKKITCECTKPDSPYPLFDGIFCSS 1706

RESULT 6  
A:45545  
major merozoite surface antigen - malaria parasite (Plasmodium falciparum) (fragments  
C:Species: Plasmodium falciparum  
C:Date: 17-Feb-1994 #sequence\_revision 17-Feb-1994 #text\_change 09-Jun-2000  
C:Accession: A45545  
R:Blackman, M.J.; Ling, I.T.; Nicholls, S.C.; Holder, A.A.  
Mol. Biochem. Parasitol. 49, 29-33, 1991  
A:Title: Proteolytic processing of the Plasmodium falciparum merozoite surface protei  
A:Reference number: A45545; MUID:92131048; PMID:1775158  
A:Accession: A45545  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-400 <BLA>  
A:Note: sequence extracted from NCBI backbone (NCBIN:77612, NCBI:77621)  
C:Superfamily: major merozoite surface antigen  
C:Keywords: glycoprotein; merozoite; surface antigen; tandem repeat

Query Match 51.6%; Score 1175; DB 2; Length 400;  
Best Local Similarity 56.6%; Pred. No. 2.8e-53;  
Matches 226; Conservative 59; Mismatches 90; Indels 24; Gaps 4;

QY 34 PDLGTDDDKAMADIGSIEGRGTMAISVTMDNLSGFENEYDVIYKPLAGVYRSLLKQI 93  
Db 5 PIFGESEED--YDDLQGVVTGEAVTPSV-IDNLSIKIENEYEVLYLKLPLAGVYRSLLKQ 61  
QY 94 EKNIFTNENLNDILNSRLKRRKYFDLVLESQKPKHISSEYIETDSFKLLNSQKNT 153  
Db 62 ENNVTFNVVNDILNSRKNRKNFRNVLDESDDIPIKDTSSVNVKDPYKFLNKKERDK 121  
QY 154 LUKSYKIKESVENDIKFAQEGISYEKVLAKYKDDLESIKKVIKEEKEFPSSPTTP 213  
Db 122 FLSSYNYIKDSIDTDINFANDVLGYKYLSEKYSKLDLSIKYI----- 165  
QY 214 SPAKTDEQKESKFLPFLNIETLYNNLVNKIDYILINLAKKINDCNVKEDEAHVKITKL 273  
Db 166 ---NDKQGENEKYLPFLNNIETLYKTVANDKIDLFVHLEAKVLYNTEKSNVEVRIKEL 221  
QY 274 SLDKAITDDKIDLFKNPYDEAIAKKLNDOTKKDMLGKLLSTGLV-QNFPNTIISKLEGG 332  
Db 222 NYLKTIDKLADFKNNNFVGIADLSTDYNNHNLTKFLSTGWFENLAKTVLSNLLDGN 281  
QY 333 FODMLNISQHCVKKQCPENSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENNGG 392  
Db 282 LQMLNISQHCVKKQCPENSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENNGG 341  
QY 393 CDADATCTEEDSGSSRKITCECTKPDSPYPLFDGIFCSS 431  
Db 342 CDADAKTEEDSGSGNGKKITCECTKPDSPYPLFDGIFCSS 380

RESULT 7  
S05603  
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)  
N:Alternate names: gp195 surface antigen  
C:Species: Plasmodium falciparum  
C:Date: 12-Feb-1993 #sequence\_revision 12-Feb-1993 #text\_change 09-Jun-2000  
C:Accession: S05603; S04850  
R:Myler, P.J.  
submitted to the EMBL Data Library, April 1989  
A:Reference number: S05603  
A:Accession: S05603  
A:Molecule type: mRNA  
A:Residues: 1-1639 <MYL>  
A:Cross-references: EMBL:X15063; NID:g9896; PIDN:CAA33163.1; PID:g9897  
Nucleic Acids Res. 17, 5401, 1989  
A:Title: Nucleotide and deduced amino acid sequence of the gp195 (MSA-1) gene from pl  
A:Reference number: S04850; MUID:89345116; PMID:2668887  
A:Accession: S04850  
A:Molecule type: mRNA  
A:Residues: 1504-1639 <MYL2>  
A:Cross-references: EMBL:X15063

C:Superfamily: major merozoite surface antigen  
C:Keywords: glycoprotein; merozoite; surface antigen  
F;1-19/Domain: signal sequence #status predicted <SIG>  
F;20-1639/Product: major merozoite surface antigen #status predicted <MAT>

Query Match 51.6%; Score 1175; DB 2; Length 1639;  
Best Local Similarity 56.6%; Pred. No. 1.4e-52;  
Matches 226; Conservative 59; Mismatches 90; Indels 24; Gaps 4;

QY 34 PDLTDDDDKAMADIGSIEGRGTMAISVTMDNILSGFENEYDVLYLPLAGVYRSKKQI 93  
DB 1244 PIGSEED--YDDLQGVVTGEAVTPSV-IDNILSKTIENYEYVLYLPLAGVYRSKKQL 1300  
QY 94 EKNIFTNLNLNDILNSRLKRRKRYFLDVLSDLMQFHHISSNEYIIDSFKLLNSEQKNT 153  
DB 1301 ENNVMTENVVVKDILNSRKNFKNFVLESLLIPYKDLTSSNVVVKDPYKFLNKKRDK 1360  
QY 154 LLKSYKIKESVENDIKFAOEGISYYEKLAKYKDDLESIKKVIKEKEKFPSPPTTPP 213  
DB 1361 FLSSYNYKIDSIDTDFINFANDVLGYKILSEKYSKDLSDSIKKYI----- 1404  
QY 214 SPKATDQKKESFLPFLTNIETLYNNLVNKIDYTLKAKINDCNVKEDEAHVKITKL 273  
DB 1405 ----NDKQGENEYKLPFLNIETLYKTVNDKIDFLVHLEAKVLNYTEKSNVEVKIKEL 1460  
QY 274 SOLKAIDDKIDLFPKNPYDFPAIKKLLINDDTKKOMGLKLLSTGLV-QNFPNTIISKLEGG 332  
DB 1461 NYLKTIOKDLADFKNNFVGIADLSTDYNNHNLTKFLSTGMVFENLAKTVLSNLLDGN 1520  
QY 333 FQMLNISQHCYKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNTCENNGG 392  
DB 1521 LQGLNISQHCYKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNTCENNGG 1580  
QY 393 CDADATCTEDSGSSRRKITCECTKPDSPYPLFDGIFCSS 431  
DB 1581 CDADAKCTEDSGSGNKKITCECTKPDSPYPLFDGIFCSS 1619

RESULT 8  
SAZQK1  
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (st  
C:Species: Plasmodium falciparum  
C:Date: 30-Sep-1987 #sequence\_revision 30-Sep-1987 #text\_change 09-Jun-2000  
C:Accession: A25120  
R;Mackay, M.; Goman, M.; Hyde, J.E.; Scaife, J.; Certa, U.; Stunnenberg, H.; H  
EMBO J. 4, 3823-3829, 1985  
A;Title: Polymorphism of the precursor for the major surface antigens of Plasmodium fal  
A;Reference number: A91030; MUID:86136024; PMID:3004972  
A;Accession: A25120  
A;Molecule type: DNA  
A;Residues: 1-1631 <MAC>  
C;Comment: The merozoite stages of different strains have strain-specific surface antige  
C;Superfamily: major merozoite surface antigen  
C:Keywords: glycoprotein; merozoite; surface antigen; tandem repeat; transmembrane prote  
F;1-19/Domain: signal sequence #status predicted <SIG>  
F;20-1631/Product: major merozoite surface antigen #status predicted <MAT>  
F;67-84/Region: 3-residue repeats (S-G-T/p)  
F;1614-1631/Domain: membrane anchor #status predicted <MBN>  
F;97,259,755,759,835,911,955,1049,1156,1165,1436,1563/Binding site: carbohydrate (Asn) (

Query Match 46.2%; Score 1051.5; DB 1; Length 1631;  
Best Local Similarity 52.3%; Pred. No. 3e-46;  
Matches 210; Conservative 62; Mismatches 103; Indels 25; Gaps 5;

QY 34 PDLTDDDDKAMADIGSIEGRGTMAISVTMDNILSGFENEYDVLYLPLAGVYRSKKQI 93  
DB 1235 PIGSEED--YDDLQGVVTGEAVTPSV-IDNILSKTIENYEYVLYLPLAGVYRSKKQL 1291  
QY 94 EKNIFTNLNLNDILNSRLKRRKRYFLDVLSDLMQFHHISSNEYIIDSFKLLNSEQKNT 153  
DB 1292 ENNVMTENVVVKDILNSRKNFKNFVLESLLIPYKDLTSSNVVVKDPYKFLNKKRDK 1351

QY 154 LLKSYKIKESVENDIKFAOEGISYYEKLAKYKDDLESIKKVIKEKEKFPSPPTTPP 213  
DB 1352 FLSSYNYKIDSIDTDFINFANDVLGYKILSEKYSKDLSDSIKKYI----- 1395  
QY 214 SPKATDQKKESFLPFLTNIETLYNNLVNKIDYTLKAKINDCNVKEDEAHVKITKL 273  
DB 1396 ----NDKQGENEYKLPFLNIETLYKTVNDKIDFLVHLEAKVLNYTEKSNVEVKIKEL 1451  
QY 274 SDKAIDDKIDLFPKNPYDFPAIKKLLINDDTKKOMGLKLLSTGLV-QNFPNTIISKLEGG 332  
DB 1452 IYKLTIOKDLADFKNNFVGIADLSTDYNNHNLTKFLSTGMVFENLAKTVLSNLLDGN 1511  
QY 333 FQMLNISQHCYKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNTCENNGG 391  
DB 1512 LARYVKHFTTPMRKKTMIQSSGCFRHLDERECKCLLNYKQEGDKCVENPNTCENNGG 1571  
QY 392 GCADATCTEDSGSSRRKITCECTKPDSPYPLFDGIFCSS 431  
DB 1572 GCDADAKCTEDSGSGNKKITCECTKPDSPYPLFDGIFCSS 1611

RESULT 9  
A24594  
probable major surface antigen (83K, 19K, 42K) precursor - malaria parasite (Plasmodi  
C:Species: Plasmodium falciparum  
C:Date: 29-Aug-1987 #sequence\_revision 29-Aug-1987 #text\_change 09-Jun-2000  
C:Accession: A24594  
R;Holder, A.A.; Lockyer, M.J.; Odink, K.G.; Sandhu, J.S.; Riveros-Moreno, V.; Nicholl  
Nature 317, 270-273, 1985  
A;Title: Primary structure of the precursor to the three major surface antigens of Pl  
A;Reference number: A24594; MUID:86014355; PMID:2995820  
A;Accession: A24594  
A;Molecule type: DNA  
A;Residues: 1-1640 <HOL>  
C:Superfamily: major merozoite surface antigen  
C:Keywords: surface antigen

Query Match 44.3%; Score 1007.5; DB 2; Length 1640;  
Best Local Similarity 50.9%; Pred. No. 5.3e-44;  
Matches 202; Conservative 65; Mismatches 105; Indels 25; Gaps 5;

QY 37 GTDDDDKAMADIGSIEGRGTMAISVTMDNILSGFENEYDVLYLPLAGVYRSKKQIEKN 96  
DB 1247 GESEED--YDDLQGVVTGEAVTPSV-IDNILSKTIENYEYVLYLPLAGVYRSKKQLENN 1303  
QY 97 IFTFNLNLNDILNSRLKRRKRYFLDVLSDLMQFHHISSNEYIIDSFKLLNSEQKNTLLK 156  
DB 1304 VMTFNVNVVKDILNSRKNFKNFVLESLLIPYKDLTSSNVVVKDPYKFLNKKRDKFELS 1363  
QY 157 SYKIKESVENDIKFAOEGISYYEKLAKYKDDLESIKKVIKEKEKFPSPPTTPPSPA 216  
DB 1364 SYNYKIDSIDTDFINFANDVLGYKILSEKYSKDLSDSIKKYI----- 1404  
QY 217 KTDQKKESFLPFLTNIETLYNNLVNKIDYTLKAKINDCNVKEDEAHVKITKLSDL 276  
DB 1405 -NDKQGENEYKLPFLNIETLYKTVNDKIDFLVHLEAKVLNYTEKSNVEVKIKELNYL 1463  
QY 277 KAIDDKIDLFPKNPYDFPAIKKLLINDDTKKOMGLKLLSTGLV-QNFPNTIISKLEGGF-Q 334  
DB 1464 KTIOKDLADFKNNFVGIADLSTDYNNHNLTKFLSTGMVFENLAKTVLSNLLDGNLKLAR 1523  
QY 335 DMLNISQHCYKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNTCENNGGCD 394  
DB 1524 YVXHFTTPMRKKTMIQSSGCFRHLDERECKCLLNYKQEGDKCVENPNTCENNGGCD 1583  
QY 395 ADATCTEDSGSSRRKITCECTKPDSPYPLFDGIFCSS 431  
DB 1584 ADAKCTEDSGSGNKKITCECTKPDSPYPLSVIFCSS 1620

RESULT 10  
A39401  
merozoite surface antigen 1 precursor - Plasmodium vivax

C:Species: Plasmodium vivax  
C:Date: 28-Feb-1992 #sequence\_revision 28-Feb-1992 #text\_change 24-Nov-1999  
C:Accession: A39401  
R:del Portillo, H.A.; Longacre, S.; Khouri, E.; David, P.H.  
Proc. Natl. Acad. Sci. U.S.A. 88, 4030-4034, 1991  
A:Title: Primary structure of the merozoite surface antigen 1 of Plasmodium vivax reveal  
A:Reference number: A39401; MUID:91219506; PMID:2023952  
A:Accession: A39401  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1726 <DEL>  
A:Cross-references: GB:M60807  
C:Superfamily: major merozoite surface antigen  
C:Keywords: surface antigen

Query Match 35.9%; Score 816; DB 2; Length 1726;  
Best Local Similarity 39.8%; Pred. No. 3.5e-34;  
Matches 177; Conservative 87; Mismatches 141; Indels 40; Gaps 10;

QY 8 SGLVPRGCMETAAKFERQHM-----SPDLGTDGDDKAMADIGSIEGRGTM 57  
DB 1280 TGVVVP---GVVSEAAQTAKAQDYAEDYDKVIELPFGNDDGDEEDQVTTGE----- 1331  
QY 58 AIVTMDNLSGFENEYDVIYKPLAGVYRSLLKQIEKNIFTNINLNDILNSRLKRY 117  
DB 1332 ASEAPEILVPAGISDYDVVYKPLAGMYKIKKQLENHVNAFNTITDMLDSRLKRY 1391  
QY 118 FLDVLESILMOPKHISSNEYIIESDFKLLNSPQKNTLLKSYKIKESVENDIKFAOEGIS 177  
DB 1392 FLEVLNSDLNPKYSPSGEYIIPKYLIDLEKKKLLGSKYKIGASIDKDLCTANDGVN 1451  
QY 178 YIEKVLAKYKDDL-----ESIKV---IKEEKEKFPSSPTTTPSPAKTDEQKES---KF 227  
DB 1452 YINMGELYKTHUTAVNEEVKKVEADIKAEADKKIKIGSDSTKTEKTSMAKAELEY 1511  
QY 228 LPLNTIETLYNNLVNKIDYLLINKAKINDCNVEKDEAHVKITKLSLKAIDDKLDLEK 287  
DB 1512 LPLNSLQKEYESLSKVNTYDNLKVINNCOLEKAEITVKKLQDYNNKMDERLEEK 1571  
QY 288 NPYDFEAIKLLINDDTKMDLGLKLLSTGLV-QNFPNTIISKLEIEGKFDMLNISQHCYK 346  
DB 1572 -----KSEKK--NEVKSGLLEKLMKSLIKENESKILSOLLNVQTLTMSSEHTCID 1624  
QY 347 KOCPSNGCFRHLDEREECKLLNYKQEGDKCVENPNTCNENNGGCDADATCTEDSGS 406  
DB 1625 TNVPDAACRYLDGMEEWRCLLTFAEKGKCVPGSNVTCKDNNGGCAPEAECKMTDS-- 1682  
QY 407 SRKKITCECTKPDSPYLPFGIFCSS 431  
DB 1683 --NKIVCKCTKEGSEPLFEGVFCSS 1705

RESULT 11  
A45604  
major blood-stage surface antigen Pv200 - Plasmodium vivax  
C:Species: Plasmodium vivax  
C:Date: 22-Apr-1993 #sequence\_revision 18-Nov-1994 #text\_change 24-Nov-1999  
C:Accession: A45604  
R:Gibson, H.L.; Tucker, J.E.; Kaslow, D.C.; Krettli, A.U.; Collins, W.E.; Klefer, M.C.;  
Mol. Biochem. Parasitol. 50, 325-333, 1992  
A:Title: Structure and expression of the gene for Pv200, a major blood-stage surface antigen  
A:Reference number: A45604; MUID:92158013; PMID:1371329  
A:Accession: A45604  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1751 <GIB>  
A:Cross-references: GB:M75674; NID:g160608; PID:9457336  
A:Note: sequence extracted from NCBI backbone (NCBIN:83591, NCBI:83592)  
C:Superfamily: major merozoite surface antigen  
C:Keywords: surface antigen

Query Match 34.9%; Score 793.5; DB 2; Length 1751;  
Best Local Similarity 38.1%; Pred. No. 5.1e-33;

Matches 174; Conservative 87; Mismatches 135; Indels 61; Gaps 10;  
QY 9 SGLVP-RGSG-----MKETAARFERQHM-----SPDLGTDGDDKAMADIGSIEGRGT 56  
DB 1301 AGSVPATGGAAGSTEENVAQAQDYAEDYKVIAPLFGNDDGDEEDQVTTGE----- 1356  
QY 57 MAISVTMDNLSGFENEYDVIYKPLAGVYRSLLKQIEKNIFTNINLNDILNSRLKRY 116  
DB 1357 -AESEAPEILVPAGISDYDVVYKPLAGMYKTIKKOLENHVNAFNTITDMLDSRLKRN 1415  
QY 117 YFLDVLESILMOPKHISSNEYIIESDFKLLNSPQKNTLLKSYKIKESVENDIKFAOEGI 176  
DB 1416 YFLEVLNSDLNPKYSSSGEYIIPKYLIDLEKKKLLGSKYKIGASIDKDLCTANDGV 1475  
QY 177 SYIEKVLAKYKDDLESIKKVIKEEKEKFPSSPTTTPSPAKTDEQKES----- 224  
DB 1476 TYNNMGELYKTHLDGKVEIKKVEDDI-----KKODELKLGNVNSQDSKKN 1524  
QY 225 -----SKFLPFLNTIETLYNNLVNKIDYLLINKAKINDCNVEKDEAHVKITKLS 275  
DB 1525 EPIAKAELEKYLPLFNSLQKEYESLSKVNTYDNLKVINNCOLEKAEITVKKLQD 1584  
QY 276 LKAIDDKIDLEKNPYDEFAIKKLLINDTKMDLGLKLLSTGLV-QNFPNTIISKLEIEK 334  
DB 1585 YNKMDKLEEK-----KSEKK--NEVKSGLLEKLMKSLIKENESKILSOLLNVQ 1637  
QY 335 DMLNISQHCYKQKCPNSGCFRHLDEREECKLLNYKQEGDKCVENPNTCNENNGGCD 394  
DB 1638 LLTMSSEHTCIDTNVPDAACRYLDGTEWRCLLTFAEKGKCVPGSNVTCKDNNGGCA 1697  
QY 395 ADATCTEEDSGSRKKITCECTKPDSPYLPFGIFCSS 431  
DB 1698 PEAECKMTDS---NKIVCKCTKEGSEPLFEGVFCSS 1730

## RESULT 12

A28121

major merozoite surface antigen - Plasmodium yoelii (fragment)

C:Species: Plasmodium yoelii

C:Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 26-Aug-1999

C:Accession: A28121

R:Burns Jr., J.M.; Daly, T.M.; Vaidya, A.B.; Long, C.A.

Proc. Natl. Acad. Sci. U.S.A. 85, 602-606, 1988

A:Title: The 3' portion of the gene for a Plasmodium yoelii merozoite surface antigen

A:Reference number: A28121; MUID:88124889; PMID:2448778

A:Accession: A28121

A:Molecule type: DNA

A:Residues: 1-680 &lt;BUR&gt;

A:Cross-references: GB:J03612; NID:g160678; PID:g160679

A:Experimental source: strain 17XL

A:Note: the authors translated the codon GTA for residue 429 as Leu

C:Superfamily: major merozoite surface antigen

C:Keywords: surface antigen

## Query Match

24.5%; Score 556.5; DB 2; Length 680;

Best Local Similarity 31.5%; Pred. No. 2.3e-21;

Matches 135; Conservative 86; Mismatches 164; Indels 43; Gaps 11;

QY 10 GLVPRGSGMKETAARFERQHMSPDLGTDGDDKAMADIGSIEGRGTMAISVTMDNLSG 69  
DB 270 GAVP-CGSDTRVAG-----SSVDNEDDDIYQIAS-GQSE-----DAPEKDLISE 313  
QY 70 FENEYDVIYKPLAGVYRSLLKQIEKNIFTNINLNDILNSRLKRYFLDVLESILMQF 129  
DB 314 FTNESLYVYTKRLGSTYKSLKHHMLREFSTIKEDMTNGLNKKSKQRNDFLEVLSHELD 373  
QY 130 KHISSENYIIESDFKLLNSPQKNTLLKSYKIKESVENDIKFAOEGISYEKVLAKYKDD 189  
DB 374 KDLSTNKYIRNPYQLDNDKKDKQIVNLKYATKGINEDIETTTDGIKFFKNKVEVNTQ 433  
QY 190 LESIKKVIKEEKEKFPSSPTTTPSPAKTDEQKES--SKFLPFLNTIETLYNNLVNKID 247  
DB 434 LAVKEQIATIE-----AETNDTKEEKKRYPILEDLKLGYETVIGQAE 479

[illegible][illegible]

RESULT 15  
E71606  
hypothetical protein PFB0765w - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 13-Nov-1998 #sequence\_version 13-Nov-1998 #text\_change 21-Jul-2000  
C:Accession: E71606  
R: Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.;  
Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.  
Science 282, 1126-1132, 1998

A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.

A:Reference number: A71600; MUID:99021743; PMID:9804551

A:Accession: E71606

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-980 <GAR>

A:Cross-references: GB:AE001417; GB:AE001362; NID:g3845271; PID:g384527

A:Experimental source: clone 3D7

C:Genetics:

A:Gene: PFB0765w

C:Superfamily: hypothetical protein MJ1322

Query Match 7.9%; Score 180.5; DB 2; Length 980;

Best Local Similarity 23.7%; Pred. No. 0.061;

Matches 102; Conservative 77; Mismatches 140; Indels 111; Gaps 23;

QY 71 ENEYDVYLLKPLAGVYRSLLKKQIEKNI-----FTFNLNLDL-----NSRLKK 114

Db 248 QNYI-----YLKK-----EYDLNKELEKNIHKGKLEHLSHCYEENQKLNEEIKRNSFIKN 301

QY 115 RYFVLDVL---ESDLMOFKHIS-----SNEYIIEI-----SFK 144

Db 302 KORKIDLLTNIEELLKKKEINIKLMKONVINKNEQLLKDIDENKNEHVNKLONE 361

QY 145 LNSEQKNTLL-KSYKYIKESVENDIKFAQEGISYEVKVLAKYKDDLESTIK-----195

Db 362 LKRELQNKCIKSDIEFCRKEKEDKLNLEDDLEKKKCIENLKDELINIKKMKWDMHM 421

QY 196 ----VIKEEKEFPSPPTTPSPAKTDEQ-----KESKFLPFLNIETLYN-NLVNKI 245

Db 422 TNEMDLSNKNVEELNINRYKNIYVELNNELDVIRKKLNDEEFLKEEKKKIDMVYKI 481

QY 246 DYDLNLKAKINDC-NVEKDEAHVITKLSDLKAIDDKIDLFKNPYDFAIKKLINDDTK 304

Db 482 KEVIOIKEKENEDISLKNQNLHLVKNEL---NEKEIILKNKYDKE-INMIIQYKN 537

QY 305 -----KDMGLKLLSTGLVGNFPNTIISKIEGKFQ-----DMLNISHOVKKOCPE 351

Db 538 KIOEERDMLNNKTKS-MDQTHKNOIBEMOEENKELRLKNVCDM-NLOSQILIK-----E 591

QY 352 NSGCFRHLDER-BECKLLNLYK-----OEQDKCVENPNPTCNENNGCCDADATCEE 402

Db 592 NE---RHMOEKVEEYNLLKQDQELKNIIOEYDERIEIQNKEMEDIVNDCEKL-----K 644

QY 403 DSGSSRKKT 412

Db 645 QAKINKKLT 654

RESULT 16

T28677

rhophtry protein - plasmodium yoelii

C:Species: Plasmodium yoelii

C:Accession: T28677; C45521

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 01-Dec-2000

R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.

Mol. Biochem. Parasitol. 42, 241-246, 1990

A:Title: A gene coding for a high molecular mass rhophtry protein of Plasmodium yoelii.

A:Reference number: 220508; MUID:95021522; PMID:7935623

A:Accession: T28677

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-2269 <KEE>

A:Cross-references: EMBL:L27838; NID:g457145; PID:g457146; PIDN:AAA21304.1

R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.

Mol. Biochem. Parasitol. 42, 241-246, 1990

A:Title: Identification of the gene for a plasmodium yoelii rhophtry protein. Multiple co

A:Reference number: A45521; MUID:91101660; PMID:2270106

A:Accession: C45521

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 2131-2269 <KE2>

A:Cross-references: GB:M34283

Query Match

Best Local Similarity 21.4%; Score 167.5; DB 2; Length 2269;

Matches 110; Conservative 69; Mismatches 177; Indels 157; Gaps 21;

QY 25 KFERQHMDSPDLTDDDDKAMADIGSIEGRGTMAISVTMDNIIISGFENYDVYIYKPLAG 84

Db 599 KEKLKHYDFDFGKEGNIKYTKIKIND-DINAVSQIDQIHGLD-----IQKSES 652

QY 85 VYRSLAKQIEKNFTFNLDNLILNLRKRYFLDVLSDLMQFKHISNEYIIESFK 144

Db 653 YVSEMKEQINKLEKVSNTISNDNVEGIRKKQII-----VTKIDKKKNYIEIRNK 703

QY 145 LLNS-----EQNTLLKSYKVIK-----ESVENDIKFAQEGISYEVKVLAKYKD 188

Db 704 LLSEISKIEKDNSTLEKVKDINLSYGNLGNLFLEQIDEEKKKAENTI-----KSMEAYID 759

QY 189 DLSIKKVIKE-BKE---KFPSPPTTPSPAKTDEQKESKFLPFLTNIETLYNNLVNK 244

Db 760 DLNIIKKKSQEIETEMDIKMDINKMEALKISHDDDKCHDKSNKHENISDIYDKSKI 819

QY 245 IDYILNLKAKINDC-----NVEKDEAH-----VKITKLSDLKAIDDKID 284

Db 820 IQDF--SRESDIINDIRKNLQKNVSESQNHNSDINQCLNEVANIYNIILKAKIRKIIDVK 877

QY 285 LFRNPYDFAIKKLINDC-----NVEKDEAH-----VKITKLSDLKAIDDKID 302

Db 878 EYTS--EIEKNKNINDELNNSEKVIKIEGDLSEKRSKINSITLDDKDIDECIKINIV 935

QY 303 TKKMDLGLKLLSTGLVGNFPNT-----IIS-----KLIEGKFQDMLNISHOVCV-----345

Db 936 LKKNILNE--ETNITNHFNAEYKIVLSNFNINEMADNKSQVILEIKKNNNGTNDHDYN 993

QY 346 -----KKOCPEKSCFRHLDERECKCLLN-----YKQEGDKC 378

Db 994 IKELSHKOKSNGYKTEADQNKRAIQNKELFEQY--KEEVTVLNLYYAVELANKFKDKT 1051

QY 379 VENPNPTCNENNGCCDADATCTEEDSGSSRKKI 411

Db 1052 KNDKQIIEIK--DAHNYCTLE-SGKSEKMK 1080

RESULT 17

G70163

hypothetical protein BB05112 - Lyme disease spirochete

C:Species: Borrelia burgdorferi (Lyme disease spirochete)

C:Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 08-Oct-1999

C:Accession: G70163

R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; W

son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; V

; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B

Nature 390, 580-586, 1997

A:Authors: Smith, H.O.; Venter, J.C.

A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.

A:Reference number: A70100; MUID:98065943; PMID:9403685

A:Accession: G70163

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-2166 <KLE>

A:Cross-references: GB:AE001153; GB:AE000783; NID:g2698419; PIDN:AAC68976.1; PID:g26

A:Experimental source: strain B31

Query Match

Best Local Similarity 22.9%; Score 166; DB 2; Length 2166;

Matches 75; Conservative 59; Mismatches 104; Indels 90; Gaps 15;

QY 90 KKOIEKNIFTNMLN---DILNSRLKK-RYFLDVLSDLMQFKH-----ISSNEYIIEI 141

Db 1670 QKDEDKIFELNQKLEHEFTLSSNLDKVRREVMVDVLISSOKESFEGGIELINKNISEFE 1729

QY 142 SFKLLNSEQNTLLKSYKYIKESVENDIKFAQEGISYEVKVLAKYKDDLESIKVKIIEEK 201

Db 1730 KISLYRNNIETSIENEYNSFSKISDGLGLEDEL---KSLKHSTSEIETIKSLGLOEQI 1786

QY 202 EKFPSPPTTSPAKTDQKESKELPFLTNIETLYNNLVNKIDDYLLINLAKAKINDCNV 261  
Db 1787 DKFEV-----EFKNHK-----BLLEVDNNILESKILNCV 1820  
QY 262 EKDEAHVKITKLSLKAIDDKIDL--FKNPYD-----FEATKKLINDDK----- 304  
Db 1821 QFNKF---LSEIKD-NLVEYKSDLRAEFEDSYDKINFQENIENPKNKLDSELEKNNSTF 1876  
QY 305 -----KDLGKLLST-----GLVQNPNTIISKIEGFDMLNISOHQCVKKOCPPENS 353  
Db 1877 LEAYSLLKDLKLEKWLTKLKEIGLAQYKNN-----FENVNKEFYNIQK-----ETL 1922  
QY 354 GCFRHLDERECKLLNLYKQEGDKCVEN 381  
Db 1923 GIIEIFNE-----LKLQESIKSIKN 1943  
RESULT 18  
A64505  
P115 homolog - Methanococcus jannaschii  
C:Species: Methanococcus jannaschii  
C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 02-Jun-2000  
C:Accession: A64505  
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,  
raon, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.  
Science 273, 1058-1073, 1996  
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese,  
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii  
A:Reference number: A64300; MUID:96337999; PMID:8688087  
A:Accession: A64505  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1169 <BUL>  
A:Cross-references: GB:U67604; GB:L77117; NID:g1592224; PID:g1500543; TIGR:MJ1643  
A:Map position: FOR1623481-1626990  
C:Superfamily: chromosome segregation protein SMC1  
Query Match 7.1%; Score 160.5; DB 2; Length 1169;  
Best Local Similarity 24.2%; Pred. No. 0.79;  
Matches 86; Conservative 59; Mismatches 110; Indels 101; Gaps 17;  
QY 94 EKNIFTENLNILNSRLK---KRYFLDVLSDLMQFKHSSNEYIIESFKLL----- 146  
Db 85 ENNAFVNADKVGILR-RIKSGETDYLYLVKWKENDKRRKKMTKHEII--DLFRGLGLG 141  
QY 147 -NSEQKNTLLK-----SYKIKESVENDIKFAQEGISYVEKVLAK 185  
Db 142 DNVISQDGLLKIINISPIERRKIIDEISGIAEFDEKKAEEELKARELIEMIDIRISE 201  
QY 186 YKDDLESTKKVKEKEKFPSPPTTSPAKTDQKESKELPFL----- 231  
Db 202 VENNLKLLKE-KEDAERY-----IKLNEELAAKAYALILKVSYNVLLENIQ 249  
QY 232 ---TWIETLYNNLVNK---IDDYLLINLAKINDC-----NVEKDEAHVKITKL----- 273  
Db 250 NDIKNLEELKNEFLSKVREIDVEIENLKLRLNIINELNKGNEVEVLEHKSIRELVEI 309  
QY 274 -SDLKAIDDKI-DLPKNPYDFEAIKKLLINDDTKKDMLGKLLSTGLVQNPNTIIS---K 327  
Db 310 ENDKKVLDSINELKKVEIENKKEIKETQKK-----IENRDSIIEKEQOIK 359  
QY 328 LIEGKFQDMLNISQHCVKKOCPPNSGCFRHLDERECKLLNLYKQEGDKCVENPN 383  
Db 360 EIEEKIKN-LNYKER-LKEATAESESIIKHLKESE-----MEIADETAKNQ 405  
RESULT 19  
T18372  
repeat organellar protein - Plasmodium chabaudi  
C:Species: Plasmodium chabaudi

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T18372  
R:Werner, E.B.; Taylor, W.R.; Holder, A.A.  
Mol. Biochem. Parasitol. 94, 185-196, 1998  
A:Title: A Plasmodium chabaudi protein contains a repetitive region with a predicted  
A:Reference number: Z18922; MUID:98418765; PMID:9747969  
A:Accession: T18372  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1939 <WER>  
A:Cross-references: EMBL:U43145; NID:g1151157; PID:g1151158; PIDN:AAC63403.1  
Query Match 6.9%; Score 156.5; DB 2; Length 1939;  
Best Local Similarity 22.5%; Pred. No. 2.3;  
Matches 97; Conservative 57; Mismatches 141; Indels 137; Gaps 18;  
QY 71 ENEDYVILKPLAGYVRSLLKKOIE-----KNITFFNL--NLNDI--- 107  
Db 114 EKQINKEVEKELNESEFIKROMELLKELNINKENINKEIITLKREELNDIESE 173  
QY 108 -----LNSRLKRRKYFLDVLSDLMQFKHSSNEYIIESFKLLNS---EOKNTL 154  
Db 174 YTEKNKEKELNVEYTNIKMSLDKLTCEVOEKK-----DNLEKINKKVKIENNL 223  
QY 155 LKSYKIVIESVE-----NDIKFAQE--GISYVEK----- 181  
Db 224 RELKEFMKEKNEIIESLDGTINDKKNAYEKELEISFEKRRKMIEMLSDKLIEKEENFANKQ 283  
QY 182 -----VLAKYKDDLESIKKVKKEKFPSPPTTSPAKTDQKESKELPFLT 232  
Db 284 AKLEKENETIIEKL-DIESREKDPKSEKKEKFAFM--ENELNLKSDLSKNACOMEVYKL 340  
QY 233 NIETLYNNLV-----NKKIDDYLLINLAK---INDCNVEKDEAH---VKITKL----- 273  
Db 341 EIKDLSQSLVEKEREFEIKNEVDKINNKEKLLSINDKIDNTVLHSEEEKINKLLKE 400  
QY 274 --SDLKAIDDKIDLKPNPYDFEAIKKLLINDDTKKDMLGKLLSTGLVQNPNTIISKLEIG 331  
Db 401 KETELNEIHKYNL-----EITIKNELNEKEEELKKAHTVEVTNLTKRI--KLEK 453  
QY 332 FKQDMLNISQHCVKKOCPPNSGCFRHLDERECKLLNLYKQEGDKCVENPNPTNENNG 391  
Db 454 KTED-----AKEGHKNELNQLSKLNEKDN---IKNENTELNDKIS 495  
QY 392 GCDADATCTED 403  
Db 496 SLNSEVNILNKD 507  
RESULT 20  
B70356  
chromosome assembly protein homolog - Aquifex aeolicus  
C:Species: Aquifex aeolicus  
C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 02-Jun-2000  
C:Accession: B70356  
R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;  
V. Nature 392, 353-358, 1998  
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A:Reference number: A70300; MUID:98196666; PMID:9537320  
A:Accession: B70356  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1156 <AQF>  
A:Cross-references: GB:AE000699; NID:g2983236; PIDN:AAC06839.1; PID:g2983243; GB:AE00  
A:Experimental source: strain VF5  
C:Genetics:  
A:Gene: xcpC  
C:Superfamily: chromosome segregation protein SMC1  
Query Match 6.8%; Score 154.5; DB 2; Length 1156;  
Best Local Similarity 22.4%; Pred. No. 1.6;  
Matches 91; Conservative 81; Mismatches 129; Indels 105; Gaps 21;



Qy	39	DDDDKAMADIGSTEGRTMAIS--VTMDNILSGF-----ENEYDVIV-----	78
		: : : :   :   :   :   :   :   :   :   :   :   :	
Db	96	EDEEVISRKVSVDGRSIFKINGQVVRERDLKDFLAKAGIYTAYNVVYQGDIVKFLKMT	155
		: : : :   :   :   :   :   :   :   :   :   :   :	
Qy	79	-----LKPLAGV--YRSLLK-----QIEKNITFNLDILN--SRLLKKRYFLD	120
		: : : :   :   :   :   :   :   :   :   :   :   :	
Db	156	PVERKTIIEISIGIEYERKKAELEAELVELKIKEIDILILEETSNQLRKEKEKLE	215
		: : : :   :   :   :   :   :   :   :   :   :   :	
Qy	121	VLESDDLMOFHHSNEYIIEDSKLLNSQNTLLSYKYIKIESVENDIKFAOEGISYYE	180
		: :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Db	216	KFK-ELQRIKRETEAKILLKEKELL--KERERILNELSSRESLE-DITFO---IQENE	268
		: :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Qy	181	KVLAKYKDDESLLTKKVI---KEEKEFPPSPPTTPPSPAKTDEQKESFLPFLNIETL	237
		: :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Db	269	KELNERERLLKVEYNKIMPFKEVGKFTAFIEAENASERSIKEKEREKLSE--NRVKNL	326
		: :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Qy	238	YNLVNKIDDYLNLRKAKINDCNVE-----KDEAHKVTKLSOLKAIDOKIDLFPKP	289
		: : : :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Db	327	INNLLSDKE---NLREVGTQLLEKLKEEVKSLKEVEREKRLREEERERLKI----	378
		: : : :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Qy	290	YDFEAIKKLLNDTQKMDMLGKLLSTGLVQ-----NFPNTI-----ISKLI	335
		: :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Db	379	-TFDEVKVL--EBEEKLTEKLSLNKKEQELEIQIRANLNKNTERIKEDINKLISERE	435
		: :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Qy	336	MLNISQHQ-----CVKKQCPCNSCGFRHLDERECKCLLNYKQE	374
		: :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Db	436	IKEIKEQEQRKLRKAIRKK-----EEEE---LRNLTOE	466
		: :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
RESULT 21			
B97116			
Chromosome segregation SMC protein, ATPase [Imported] - Clostridium acetobutylicum			
C:	Species:	Clostridium acetobutylicum	
C:	Date:	14-Sep-2001 #sequence_revision	
C:	Accession:	B97116	
R:	Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gilman, D.A.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.		
J.	Bacteriol.	183, 4823-4838, 2001	
A:	Title:	Genome Sequence and Comparative Analysis of the Solvent-Producing Strain B97116	
A:	Reference number:	A96900; PMID:21359325; PMID:21359325	
A:	Status:	preliminary	
A:	Molecule type:	DNA	
A:	Residues:	1-1191 <CUR>	
A:	Cross-references:	GB:AE001437; PIDN:AAK7917.1; PDB:gl5024720; GSPDB:GN001437	
A:	Experimental source:	Clostridium acetobutylicum ATCC894	
C:	Genetics:		
A:	Gene:	CAC1751	

```

Qy 298 LINDTTK---DMLGKLLSTGLVN---FNTIISKLIEGKFQDMLNISOHCVKKQCP 351
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db 408 EIESNESKISIKRAGEYSKSLKINEVTNLTSEELV--KINDKISGYENQ-IRENRSK 464
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Qy 352 NSGCFRHLDEREECKCLLNYKQEGDKCVENPPTCN 387
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db 465 ISKLNRIISDEEKLRELNSK--SNKLEANKNMLIN 498
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |

RESULT 22
T28676
rhoptry protein - Plasmodium yoelii (fragment)
C:Species: Plasmodium yoelii
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000
C:Accession: T28676; A45521
R:Sinha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.
Mol. Biochem. Parasitol. 76, 329-332, 1996
A:Title: Comparison of two members of a multigene family coding for high-molecular ma
A:Reference number: 220507; MUID:97077455; PMID:8920022
A:Accession: T28676
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-2401 <SIN>
A:Cross-references: EMBL:U36927; NID:g1041784; PID:g1041785; PIDN:AAB41263.1
R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A:Title: Identification of the gene for a plasmodium yoelii rhoptry protein. Multiple
A:Reference number: A45521; MUID:91101660; PMID:2270106
A:Accession: A45521
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 2260-2401 <KE>
A:Cross-references: GB:M34281

Query Match      6.7%; Score 152; DB 2; Length 2401;
Best Local Similarity 21.5%; Pred. No. 4, 9;
Matches 91; Conservative 82; Mismatches 159; Indels 92; Gaps 23;

Qy 64 DNIISGFENEVDVYTLKPLAGVYRSLLKQIEKNFTFNMLNDILNSRLKRRKYFLDV-- 121
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db 285 NEIYNTIKSDPKDIYVGDIQOLYNEMFSVVGQESNIEHIEKNTEILTTLTKIDNVYNNIQN 344

Qy 122 LESDLMQ--FKHISSNEYIIE--DSFKLLNSEQKNTL--LKSYYKIKESVENDI-KFA 172
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db 345 METETVSHULNIETNNKLSLTIILDIILKIYGETINELNKTLEDFKNKEKGLSNKIDEYA 404

Qy 173 QEGIS---YYEKVL---AKYKD--DLESIK-KVIKEKEKFPSPPTTPSPAKTDQKK 223
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db 405 KENVQLNVYKSNILEIKKHYNDQINIDNIKEKAKYVDQFKEHMKTIPTNEMKYKPSI 464

Qy 224 ESKEL--PFLTNIETLYNN-----LVNKI-----DDYLINLKAIND 258
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db 465 EIKIMKDFLSKVNK-KYNDFKVYKKEVSESHNNKFTLTNKRITSEVDEEIKKYENKFD 523

Qy 259 ---CNVEKDEAHVYIKTSLDKAIDDKIDLFPKNPYDFEALKLINDTK-KDWLGKLLS 313
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db 524 SKSLINETKKSIEBEYQNIPLAKKVDYDIKVLNTN--ELITNCHNQTTLKORLQNQIK 581

Qy 314 TGLVQNFNPNTIISKLEGGFQDM-----LNISOHCVKKQ-----CPEN 352
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db 582 TIKETN---SIDKIYTDKFNILTDKKTLETKFTGLSLNNHESNNKELLTYFYDLKAN 637

Qy 353 SG-----CFRHLDEREECKCLLNYKEG---DKCVENPPTCNENNGCGDADATCTEED 403
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db 638 LGKNKENMLYKQFNEKE--KAVEDIKKKNDINKIVSNIETIITYTSYININED---TENE 692

Qy 404 SGSS 407
      |
Db 693 IGS 696

RESULT 23

```

B59102  
hypothetical protein pX01-90 - Bacillus anthracis virulence plasmid pX01  
C:Species: Bacillus anthracis  
C:Date: 12-Nov-1999 #sequence\_revision 12-Nov-1999 #text\_change 11-May-2000  
C:Accession: B59102  
R:Okimaka, R.T.; Cloud, K.; Hampton, O.; Hoffmaster, A.R.; Hill, K.K.; Keim, P.; Koehler  
J. Bacteriol. 181, 6509-6515, 1999  
A:Title: Sequence and organization of pX01, the large Bacillus anthracis plasmid harbored  
A:Reference number: A59091; MUID:99445483; PMID:10515943  
A:Accession: B59102  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-652 <OKI>  
A:Cross-references: GB:AF065404; NID:G4894216; PIDN:AA032394.1; PID:G4894306  
A:Experimental source: strain Sterne  
A:Note: similar to hypothetical protein; PFB0765w (980 aa); Plasmodium falciparum (AE001  
C:Genetics:  
A:Gene: pX01-90  
A:Genome: plasmid

Query Match 6.6%; Score 149.5; DB 2; Length 652;  
Best Local Similarity 21.2%; Pred. No. 1.5;  
Matches 83; Conservative 70; Mismatches 132; Indels 107; Gaps 15;  
QY 18 MKETAARAFERQHMDSPLDGTDDDDKAMADIGSIEGRGTMAISVTMDNILSGFP----- 71  
DB 284 LKESQSRLELNKDKSNRLNLSNKKLND-----RKAELLSLIMELIKQOSEFDRKKIK 337  
QY 72 NEYDVI-----YKPLAGVYRSUKKQIEKNFTFN--NLNDILNSLRKKRYFLDVLES 124  
DB 338 NEKDLNKKREDLINRAESKELAKKAELNTKLVELFKVOEALNKKSGQVLYIINKLDN 397  
QY 125 DLMOFKHISNEYIIEFSKLLNSEQNTLKSY--RYIK--ESVENDIRFAQEGISYYE 180  
DB 398 ELRE-----LADKYK--NSDNISRLKNHIGEYNKQLEKIELECECNKKIDNTK 445  
QY 181 KVLN-----KYKDDLE-SIKKVIKEKEKFPSPSP 209  
DB 446 KQAEFDKSNKQOESELSVLQNKKIDELGKRHKHROELESASQKALDEAKENKKLAE 505  
QY 210 TTPSPAKTDEOKESKFLPPLTNIYLNVLNKKIDYILINKAKI-----NDC 259  
DB 506 KDPERHINDKLLK-----LNDRWLTNNELNRLNSELQKNTKLLGYSOEHOYNNAL 558  
QY 260 NVEKDEAHVKITLSDUKAIDDK-----IDLKNPYDFEATKILINDDTKMDLGLK 311  
DB 559 NKRETLSSKKVDEYQKQSEIDKKNYIELNQLDELSPKPKNEQVQLL-----KNQINEL 612  
QY 312 LSTGLVQNFPTTIISKLEGGFQDMLNLSQHO 343  
DB 613 -----NSSINQILNHRKKNINNTORK 634

RESULT 24  
C97325  
hypothetical protein CAC3461 [imported] - Clostridium acetobutylicum  
C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
C:Accession: C97325  
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
J. Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cl  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: C97325  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-909 <KUR>  
A:Cross-references: GB:AE001437; PIDN:AAK81390.1; PID:G15026552; GSPDB:GN00168  
A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC3461

Query Match 6.5%; Score 147; DB 2; Length 909;  
Best Local Similarity 20.8%; Pred. No. 2.9;  
Matches 99; Conservative 80; Mismatches 152; Indels 144; Gaps 22;  
QY 19 KETAARAFERQHMDSPLDGTDDDDKAMADIGSIEGRGTMAISVTMDNILSGFENEYDVIY 78  
DB 358 KENEIKRIAKSAMEDVNSMDSDLNALISELYNIESR-----MDQIQRNLKLLKDLVE 409  
QY 79 LKPLAGVYRSUK-----KQIEKNFTFN----- 101  
DB 410 LKKEAQLENAKEKLDILEYKPIWVTRNKHKNELFKNSYERKYEKEIKEYNOAEDELK 469  
QY 102 -NLNDILNSLRKKRYFLDVLESQMFKHIS-----SNEYIIE-----DSFKLL 146  
DB 470 SMNISNL--DQYDKAKIIFDDKEQELIEAENSIASLDAEKENSIIKYKAKRDDIIKQ 527  
QY 147 NSEQNTLLASIKYKIKESVENDIKFAQEGISYYEKVLAK-----YKDDLESIKKVIKE 199  
DB 528 NSDILNVNMTY-FADEVKEDSSKKIQTIN--EKELKKKDEINKENELKENKNDIVKE 584  
QY 200 E-----KEKFPSPPTTPSPAKTDEQKESK--ELPFLTNIETLYNINLVNKID----- 246  
DB 585 ESLPKENNVNKKQEDNTKVPADNKNKCKQYKKNKNNKSVKNGOSKNNTOKNPKOKKTL 644  
QY 247 -----DYLINLAK-----INDCNVEKDE---AHVKITKL-SDLK----- 277  
DB 645 EQLKELLCSLAKAKDELKISKNTISEDMAFCEKVKKIDEDIKILDKETISGLVGLKS 704  
QY 278 AIDDKIDLFK-----NPYDFEAIKLLINDDTKMDLGLKLLSTGLVQNFPTTIISK 327  
DB 705 EIDDKINMKEKFFVSKMYNPIQERYSRKN-SLRKNOIREITERAMAN-----MDK 757  
QY 328 LIEGKFOOM--LNISOHCQVKKQCPENSGCFRHLDERECKCLLNYKQEGDKVC 380  
DB 758 AIVEEFDKISKLHTIENRIVEVQ-----KVLDLAQQ---MIDLEEGRKLKEE 801

RESULT 25  
S51972  
SPC72 protein - yeast (Saccharomyces cerevisiae)  
N:Alternate names: FUN42 protein; protein YAL047c; STU2 protein interactant  
C:Species: Saccharomyces cerevisiae  
C:Date: 05-May-1995 #sequence\_revision 01-Sep-1995 #text\_change 19-Apr-2002  
C:Accession: S51972  
R:Bussey, H.; Kaback, D.B.; Zhong, W.; Vo, D.T.; Clark, M.W.; Fortin, N.; Hall, J.; O  
submitted to the EMBL Data Library, August 1994  
A:Description: The sequence of chromosome 1 of Saccharomyces cerevisiae.  
A:Reference number: S51956  
A:Accession: S51972  
A:Molecule type: DNA  
A:Residues: 1-622 <BUS>  
A:Cross-references: EMBL:U12980; NID:G1326053; PID:G595537; GSPDB:GN00001; MIPS:YAL04  
C:Genetics:  
A:Gene: SGD:SPC72; SPC72; FUN42; SPI6; MIPS:YAL047C  
A:Cross-references: MIPS:YAL047c; SGD:S0000045  
A:Map position: 1L

Query Match 6.4%; Score 146.5; DB 2; Length 622;  
Best Local Similarity 18.7%; Pred. No. 2;  
Matches 80; Conservative 75; Mismatches 125; Indels 147; Gaps 18;  
QY 18 MKETAARAFERQHMDSPLDGTDDDDKAMADIGSIEGRGTMAISVTMDNILSGFENEYDVIY 54  
DB 228 LAQSSPAGSLESRSPSKEENTDGGYQNDIHSNNHIDTENVMANSTSLPSAVESR 287  
QY 55 GTMAISVTMDNILSGFENEYDVIYKPLAGVYRSUKKQIEKNFTFNLDILNSLRKK 114  
DB 288 FEKTLDTQLEIYELHKEVDQF-----INSIRLKFEK-----SOKLEKIATSLNE 334  
QY 115 RYKFLDVLESQMFKHISNEYIIEFSKLLNSEQNTLKSYKIKESVENDIKFAQ 174  
DB 335 QSHLLDSLELEENSSSVIEKQDHLISQLKEI--ESQSVLNNLEKLKEDI--IKMKQN 389

```

Db      818 DE1CEKQNG-----EG1IC 832
      | | : : ||| |
RESULT 27
T18427
hypothetical protein C0335c : malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T18427
R:Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A:Reference number: Z18935
A:Accession: T18427
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-3724 <LAW>
A:Cross-references: EMBL:Z98547; NID:el325376; PID:el325379; PIDN:CAB11104.
C:Genetics:
A:Introns: 307/1; 1545/2
A:Note: C0335c

```

```

Query Match          6.4%; Score 146; DB 2; Length 3724;
Best Local Similarity 20.4%; Pred. No. 17;
Matches 98; Conservative 65; Mismatches 148; Indels 170; Gaps

QY  2 HHHHHHSSGLVPRGSMKTAARQHUMDS-----PDL---GTTDDDKAMADIGSIEGR 54
    |||  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
Db  2985 HTNHTVS-----KGDGRKNPNDAOSEYRNKENQIKIPNIDKKNVDNNDK----- 3029
    |||  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
QY  55 GTMAISVTMDNILS-----GFENEVDVIVLKLACGV-----RSLKKQIEKNITFF 100
    |||  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
Db  3030 -----VEKDNLDLDDIKCSLKNKKIIEYLLKYSINIWIYDIEKYHSKIKMLIEK----- 3078
    |||  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
QY  101 NLNLNDILNSRLKKRYF-LDVLESDLMOFKHISSENYIIEDSEFKLLNSEQNTLLKSYK 159
    |||  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
Db  3079 --RLNDIL--RYRQNLVHQLNMKRDKICAQFLRDNVYKTSNF-----F---YF 3120
    |||  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
QY  160 YIKESVENDIKFAQEGISYIEYKVLAKYKDDLESIKKVIKEEKFPSSPPT-----TPPS 214
    |||  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
Db  3121 YCPVDQDNVIDDTEFLILYNYLFYFR-TLEQINEFVKN-----PTFYLKQLKPN 3170
    |||  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
QY  215 PAKTDEQKESKELPEL-----TNIETLYNNL-----VNKIDDYLLNL 252
    |||  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
Db  3171 P-----SFLVPFTFVYSCIEIENLHNEUHKTKCVINREHLLSFGNKKIKKYEKI 3221
    |||  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
QY  253 KAKINDCNVEKDAHVKKITKLSDLKAIDDKIDLFKNPYDFEAIKKLIINDDTKKDMLGKLL 312
    |||  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
Db  3222 QTNNDIDNILMEILHV--RLNQYDILSHGCTILYNIPYNNQLYFLLNNAKKKKYVPHLI 3278
    |||  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
QY  313 STGLVQNPFTIISKILEGKFQDMLNISOHQCVKQCPENSGCFRHLDERECKLLNLYK 372
    |||  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
Db  3279 F-IIRK-----HDKNKTNYEIIINY- 3297
    |||  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
QY  373 QEGDKCVEN-----PNPTCNENGGCDATCTEEDSGSSRKKITCECT-----RPDS 420
    |||  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
Db  3298 --NDKCIITDNMYKQNVNHNINSPNDKDNHIMNDEYKKKKYINTSFTLLSKKKNKEDS 3355
    |||  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
QY  421 Y 421
Db  3356 Y 3356

RESULT 28
T34081
hypothetical protein C02F12.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34081
R:Miller, N.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid C02F12.
A:Reference number: 221473

```

```

Query Match      6.4%; Score 146; DB 2: Length 1619;
Best Local Similarity 21.6%; Pred. NO. 6.4;
Matches 95; Conservative 65; Mismatches 133; Indels 146; Gaps 23;

QY      50  SIEGRGTMAISVTMDNTSLSGFENEYDVI-----YLKPLAGVY-----RSLKK--- 91
DB      481  NIKRNPINIKTTQOLKIKLDLNEIEKLRNALNLSRKRGVYLDTEYNNQNLSKKNKE 540

QY      92  ---OIEKNITFNUNLNDIINSR-----LKKRYFLDVLSDLMQKHISSEYIIEDSF 143
DB      541  ILLEKEKILFEKSKKIKNLNKNKMDTDLQNKIIPP--LKDITQYKNQIAFHQIL---I 595

QY      144  KLLNSEQKNT--LLKSVKYIKESVENDIKKFAQEGISYVEKVLAKYKDDLESIKKIVKEEK 201
DB      596  KKIIEEKYITQFLNKKYHSEENFINOKKSFQQIKTNMSNM---HDNLLNKKNIHQEN 652

QY      202  EKFPSSPPTPPSPAKT-----DBQKESKFLPFLTNIETLYNNLVNKIDYILNLKAKI 256
DB      653  HLLNDI-----CKTILFLDPTKN-----YLVONKKSYDINIHFEELNSNINFKY 698

QY      257  NDCNVKDEAHVKITKLSDLKAIIDDKIDLF--KNPYDF-----EATKK-----LIND 301
DB      699  ND-----TLNFLDHDHILQYDHDYFSLSENKKEISQCKINPHISD 742

QY      302  DTKKDM---LGKLLSTGLVQNFPNTIISKLETKFQDMLNTISQHCVKKQCPENSGCFRR 358
DB      743  EAKQKMPAIIIGVYNTNGLK-----EKSGEKIICEKQ 775

QY      359  LDREECKCLLNKQECGDKCVENPNPTCNENNGCCDADATCTEEDSGSSRRKKTICE--- 414
DB      776  SGEKIICE---KQSGEKII-----CEKONG-----EEICEKQSG---EKIICEKONG 817

QY      415  ----CTKRPDSYPLFDGIFC 429

```



```

Db      123 NIEVKKILSLEPLNIETENSDPOLKEIIG-----SUNDLINSNDNS 164
Qy      116 KYFLDVLESDLMOFKHSSNEYIIEDSFKLNSFOKWTLLSKSYKIKESVENDIKFAQEG 175
Db      165 ---LSIPSENFSSIKKQN-----MUTNLSQNETLK-----EA 196
Qy      176 ISYVEKVLAKYKDDLESIKKVKIEEKEKFPSPPTPPSPAKTDEQKKESKFLPFLTNIE 235
Db      197 LSNYM-----IQTSEVITELNTQYENS-----KREFE-----K 225
Qy      236 TLYNNLVNKIDYVLINL-----KAKINDCNVEKDEAH-----VKITKLSDL-KAID 280
Db      226 NLQKNLLQEVDEPKENLTQKQRELEBEKLANEBELLOAKHANEVGLLSITQVKEFNKIIK 285
Qy      281 DKIDLFKNP--YDPEATKILINDDTKK-DMLGKLLS--TGLVQ-----NF 320
Db      286 DKIEKRGRLAHEETNSEVNDKSIDRSKSLKNEALVOLTFQVDIKSRINNNNL 345
Qy      321 PNTIISLIEGKFQDMLNLSIQHCVKKQCPENSGCFRHLDEREECKLLNKKOEGD--- 376
Db      346 PDVNIDK--ELSRLLKLSNLLSTFNKSKCDDGDC-----CSCKGNKNEGKGI 394
Qy      377 --KCVENPNPTCENNGGCDADATCTEEDSGSRKKI 411
Db      395 SCCKPKPTNPSSLVALDELESTC-----SGKKI 424

RESULT 33
B90395
purine NTPase [Imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C:Accession: B90395
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayev,
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A:Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139
A:Accession: B90395
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-864 <KUR>
A:Cross-references: GB:AE006641; NID:g13815551; PIDN:AAK42417.1; GSPDB:GNO
C:Genetics:
A:Gene: SSO2249

Query Match      6.3%; Score 142.5; DB 2; Length 864;
Best Local Similarity 22.8%; Pred. No. 4.7;
Matches 87; Conservative 63; Mismatches 150; Indels 81; Gaps

Qy      27 ERQHMSDFDLGTDDDD-----KAMADIGIEGRGTWMAISVTMDLTLSGF 70
Db      364 ERLEKDASELSDNDIKVNSLEQVETPRKKQLNLAOLAKVE--SLISEKNEIINISQV 421
Qy      71 ENEYDVLYLPLAGVYR-SLKQTEKNIETFNLNLDILNSRLKRRKYFLDVLESDLMOF 129
Db      422 EGETPCVGRPLDEEHQKIIKAKSVILOLELKNKN-LDEELKKTITNEUNKIE---REY 477
Qy      130 KHTSSNEYIIED---SPKLLNSEQKN--TLLSKYKIKESVENDIKFAQEGISYVEK-- 182
Db      478 RRLSNKASYDVMYRQLKNEETENLHSESLKNIDEIKKINEEVKELKLYEEFMR 537
Qy      183 LAKY-KDDLESIK-----KVIEEEKK-----FPSSPPTPPSPAKTD 219
Db      538 LSKYTKTEELDKRVKLDKMKKEIEIEKMRGSELSKLDLRKALESKILDLENKRVKLD 597
Qy      220 EOKK-----ESKFLPFLNI-----ETLYNNLVNKIDYVLINLKAKIND 258
Db      598 EMKKKKGILEDYTRQVKLLQOEYKVNLRREYNIITQFDENRYNELKTSIDAYNLSLKEKN- 656
Qy      259 CNVEKDEAHVKITKL--SDLAIDDKTDLFKNPYDFFAIIKKI.INDDTKKMW/GKILSTGLV 317

```

C:Superfamily: hypothetical protein MJ1322									
Query Match		6.2%; Score 142; DB 2; Length 1005;							
Best Local Similarity		22.7%; Pred. No. 5.9;							
Matches 82; Conservative 68; Mismatches 117; Indels 94; Gaps									
Qy	63	MDNTLSGFENEYD-VIVYKPLAGVYRSLKKQIERKNITFTNMLNDILNSRLKK-----RK 116							
Db	346	LDTLNKKIDIEIRVETIKOLLDELKNLNEIEK-----IEYKRICEECKE 392							
Qy	117	YFLDVLSDLMQFHSHSSNVEITDSFKLNSQK-----NTLLSKYKVIK- 162							
Db	393	YYEYKLE---LEEKAVEYNKLTLE-YITLQEKKSIEKINLDLETRINKLLETNIDI 447							
Qy	163	ESVENDIKFAQEGISYVE-----KVLAKYKDDLESIKKIVKEEKEKFPSSPTTP 212							
Db	448	ESIENLSKEIEBKVKVLENQKEKIELNKKLGEINSEIKRLKVLDELKEVEGKCPCKT 507							
Qy	213	PSPAKTDQEKESKFLPFLNIETLYNNLVNKIDVYLINLAKKANDCNVEK-----DEA 266							
Db	508	P-----IDENKK---MELINQHKTLQANKYTELEE--INKKIREIKDIEKLAKKEIDKEE 557							
Qy	267	HVK-----ITKLSDLKAIDDKIDLFKNPYDFAIRKKLIN-----DDTKDMLGKLL 312							
Db	558	NLKLTKLTLYLEKQSOIELELKLKNYKEQLD-EINKKISNVINGKPVDEILEDIKSQL- 615							
Qy	313	STGLVQNPWTITSKL-----LEGFDQMLNISOHQCVKQCPENSGCFRHLDER 362							
Db	616	--NKFKNFYNOYSAVSYLNSVDEEGIRNRRIKETENIVS-GWNKEKCREELNKLRE-DER 671							
Qy	363	E 363							
Db	672	E 672							
RESULT 36									
Tl8355									
hypothetical protein p3 - Mycoplasma hyorhinis									
C:Species: Mycoplasma hyorhinis									
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 07-Dec-1999									
A:Accession: Tl8355									
R:Deng, G.; McIntosh, M. A.									
J. Bacteriol. 176, 5929-5937, 1994									
A:Title: An amplifiable DNA region from the Mycoplasma hyorhinis genome.									
A:Reference number: Z18888; MUID:95014025; PMID:7928953									
A:Accession: Tl8355									
A:Status: preliminary; translated from GB/EMBL/DDBJ									
A:Molecule type: DNA									
A:Residues: 1-1187 <DEN>									
A:Cross-references: EMBL:L11447; NID:g150156; PID:g664956; PIDN:AAA62228.1									
C:Genetics:									
A:Genetic code: SGC3									
Query Match		6.2%; Score 142; DB 2; Length 1187;							
Best Local Similarity		20.7%; Pred. No. 7.1;							
Matches 90; Conservative 61; Mismatches 138; Indels 146; Gaps									
Qy	74	YDVLYKPLAGVY-----RSUKKQIE-----KNITFTNMLNDILNSRLKK 114							
Db	25	YRLRYAKKHVGHIDFNFDIVVAINSLNAPARLWVREDNNFRRLRFLPNVNLKLTTE--QR 82							
Qy	115	RKYFLDVLSDLQMFQKHSSNVEITDSFKLNSQKNTLLSKYKIKESVENDIKFAQE 174							
Db	83	LEFYDNVF-SDR-----KELSNVELAEQFQELAEKNKERTLELAK----- 123							
Qy	175	GISYVEKVLAKYKDDLESIKKIVKEE-----KPPSSPTTPPSPAKTDQCKE 224							
Db	124	-IDSLKTI-QELQEDTESIIITTEEPQEEHQCEICIRLASSKSPVEATLGEVVEEQTV 181							
Qy	225	SKFLPFLTNIETLYNNLVNK-----IDYLI-----NLKAKINDCNVEKDE 265							
Db	182	SE-EPVVLKKEETVEKIVSNKECLSKULLDEVISTQVCQCANSKNVCDLCSNCQCHN 240							

Qy 90 KKQIEKNITFN-----LNLNDILNSRLKKRYFLDVLESIDLMPKHISSEYI- 138  
| | : : | : | : | : | : | : | :  
Db 333 KNSINNNLYQLNSLLKEKKNDFSLSDEN--RYTKSFTELVDLIILSVLSRAKSEFVL 389

Qy 139 ----IEDSFKILLNSEOKNTLLKSYYIKESVENDIKFAOEGTISYKERVLAKYKDDLE-- 191

Db 390 LKEHLSLKLUFERS-----LSIKYIKERIVNIKY-----ISKDENLALLRDKIEPIF 439  
Qy 192 ----SIRKVIKEEKEKFPSPPTPPSPAKTDEQKESKFLPFLTNITLYNNLVNKIDD 247  
Db 440 DONVNLKKFLEKN-----ALAKSLAKE-----ITNIERLIEEKTQLND 479  
Qy 248 YLINLAKKINDCNVEKDEAHVYIKY---LSDLKAIDDKIDLFKNPYDFEATKKLINDDTK 304  
Db 480 VLGELE-----YIELSKFNLDIKLIDGNLEFL-----FESKNSL--DEEL 519  
Qy 305 KDMGLKLLSTGLVONFPNTIISKLI-----EGKFODMLNISOHQCVKKQCPENS 353  
Db 520 KDLYLKENLNLKSDQLNLSNLIGASKFSSESFKEDFNSLVFLNDF-----KKT 571  
Qy 354 GCFRHLDERECKCLLNYKQGDCKVCNPNPTCNENGGCDADATCTEEDSGSS---RKK 410  
Db 572 NYEYVIRKQTEYEFLLNSVSIKSEIENFN-VSNEM-----SDKFELEEDLATRELLRKE 625  
Qy 411 ITCECTKPDSPYLPDGI 428  
Db 626 I-----DAIKLGDYVF 636

RESULT 40  
C89910  
hypothetical protein SAI181 [Imported] - Staphylococcus aureus (strain N315)  
C:Species: Staphylococcus aureus  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C:Accession: C89910  
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc  
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;  
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001  
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A:Reference number: A89758; MUID:21311952; PMID:11418146  
A:Accession: C89910  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1009 <KUR>  
A:Cross-references: GB:BA000018; PID:gl3701144; PIDN:BA42439.1; GSPDB:GN00149  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: SAI181

Query Match 6.28; Score 140; DB 2; Length 1009;  
Best Local Similarity 22.98; Pred. No. 7.5;  
Matches 97; Conservative 51; Mismatches 131; Indels 144; Gaps 21;

Qy 19 KETAAKFERQHMDSPDLCTDDDDKAMADIGSIEGRGTWASVTMDNI-----L 67  
Db 193 KEKAIENRYQOIDLWQIESEFDDK----IKGLLELATQOIKLIENIPLLQARSKEI 248  
Qy 68 SGFEN-----EYDVIYKPLAGVYRSKAKQIEKNIFTNLDNLNLSRLKRRYFL 119  
Db 249 LAFVNESKETAKEVEIE-----KKTLENNILKDNIN-----QLNKNKI-- 288  
Qy 120 DVLESDLMPKHSSNEYIIEDSFKLLNSEQNTLLSKYIKESVENDIKFAQEGISYY 179  
Db 289 -----DFVLKEQOPEIDEIEAKLKL---QDITNLLNYIENREKIETKIANSKKDISKT 340  
Qy 180 EKLAKY---KDDLESIKKVIKEKEKFPSPPTTPSPAKTDEQKESKFLPFLTNIET 236  
Db 341 NKLNLDCDRNTDKKMLLENGDLIES-----KTS-----FIDKTRV 380  
Qy 237 LYNNLVNKIDDYLYNLKAKINDCNVEKDEAHVITKLSDLKAIDDKIDLFKNPYDFEAIK 296  
Db 381 LFND-INKYQOYSLNEICLITEGQLGDELN---NLIKLEKVEDSIG--NNESDYEKII 434  
Qy 297 KLIN-----DDTKDKMLGKLLSTGLVQNFN-----TIISKLIEGKFQDML 337  
Db 435 ELNNAITNINNEINIENKAKAELDKLL--GSKQLENOINEETTMMKNLEIK----- 487

Qy 338 NISOHQCVKKQCPENSGCFRHLDERECKCLLNYKQE-----GDKCVENPNPTC 386  
Db 488 -----LDHYDKSKLDLNDKESFISEIKSAVKIGDQD-----PIC 521  
Qy 387 -NE 388  
Db 522 GNE 524

Search completed: May 12, 2003, 10:20:30  
Job time : 70 secs



GenCore version 5.1.4\_p5-4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 10:09:22 ; Search time 25 Seconds  
(without alignments)  
715.052 Million cell updates/sec

Title: US-10-057-531A-2  
Perfect score: 2275  
Sequence: 1 MHHHHSSGLVPRSGMKE.....TCECTKPDSPFLDGFICSS 431

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 segs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1946.5	85.6	1701	1 MSP1_PLAFF	P13819 plasmodium
2	1946.5	85.6	1701	1 MSP1_PLAFM	P08569 plasmodium
3	1935.5	85.1	1726	1 MSP1_PLAFC	P04934 plasmodium
4	1935	85.1	1682	1 MSP1_PLAF3	P19598 plasmodium
5	1932.5	84.9	1726	1 MSP1_PLAPF	P50495 plasmodium
6	1175	51.6	1630	1 MSP1_PLAPK	P04932 plasmodium
7	1175	51.6	1639	1 MSP1_PLAFW	P04933 plasmodium
8	556.5	24.5	1772	1 MSP1_PLAYO	P13828 plasmodium
9	160.5	7.1	1169	1 SMC_METJA	Q59037 methanococc
10	146.5	6.4	622	1 YAE7_YEAST	P39723 saccharomyc
11	145	6.4	1130	1 YL17_CAEEL	Q11102 caenorhabdi
12	143	6.3	540	1 YK26_YEAST	P36112 saccharomyc
13	142.5	6.3	864	1 RASO_SULSO	Q97wh0 sulfolobus
14	142	6.2	1005	1 RASO_METJA	Q58718 methanococc
15	142	6.2	1679	1 Y109_YEAST	P40457 saccharomyc
16	139.5	6.1	879	1 RASO_SULTO	Q96yrs sulfolobus
17	138.5	6.1	2116	1 MYS2_DICDI	P08799 dictyostell
18	136	6.0	1163	1 SBCC_CLOAB	Q971k1 clostridium
19	135.5	6.0	2869	1 RBP1_PLAVB	Q00798 plasmodium
20	135	5.9	880	1 RASO_PYRAB	Q9uzc8 pyrococcus
21	134	5.9	1790	1 USOL_YEAST	P25386 saccharomyc
22	134	5.9	1928	1 MYS1_YEAST	P08564 saccharomyc
23	133.5	5.9	1170	1 SMC2_YEAST	P38989 saccharomyc
24	133	5.8	978	1 RASO_AQUAE	O67124 aquifex aeo
25	133	5.8	2104	1 MYS3_SCHPO	O14157 schizosacch
26	132.5	5.8	886	1 RASO_SULAC	O33600 sulfolobus
27	132	5.8	882	1 RASO_PYRFU	P58301 pyrococcus
28	132	5.8	3329	1 BRC2_MOUSE	P97929 mus musculu
29	131.5	5.8	707	1 YJ9C_YEAST	P47166 saccharomyc
30	131.5	5.8	1225	1 SMC1_YEAST	P32908 saccharomyc
31	131	5.8	1251	1 RBP2_PLAVB	Q00799 plasmodium
32	130.5	5.7	895	1 RASO_THREVO	P08502 thermoplasm
33	130	5.7	886	1 RASO_ARCFU	O29230 archaeoglob

RESULT 1				
MSPI_PLAFF	STANDARD;	PRT;	1701 AA.	
ID	MSPI_PLAFF			
AC	P13819;			
DT	01-JAN-1990 (Rel. 13, Created)			
DT	01-JAN-1990 (Rel. 13, Last sequence update)			
DT	01-OCT-1996 (Rel. 34, Last annotation update)			
DE	Merozoite surface protein 1 precursor (Merozoite surface antigens)			
DE	(PMMSA).			
GN	MSP-1.			
OS	Plasmodium falciparum (isolate FC27 / Papua New Guinea).			
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
OX	NCBI_TaxID=5837;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=88142999; PubMed=2449612;			
RA	Peterson M.G. Coppel R.L., McIntyre P., Langford C.J., Woodrow G.,			
RA	Brown G.V., Anders R.F., Kemp D.J.;			
RT	"Variation in the precursor to the major merozoite surface antigens			
RT	of Plasmodium falciparum."			
RL	Mol. Biochem. Parasitol. 27:291-302(1988).			
CC	-!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor			
CC	(Potential).			
CC	-!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42			
CC	kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF			
CC	MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL: M19143; AAA29653.1; -			
DR	PIR: A54498; A54498.			
DR	InterPro: IPR000561; EGF-like.			
DR	Pfam: PF00008; EGF; 1.			
KW	Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;			
KW	Transmembrane; GPI-anchor.			
FT	SIGNAL 1 19			POTENTIAL
FT	CHAIN 20 1701			MEROZOITE SURFACE PROTEIN 1.
FT	CARBOHYD 110 110			N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 239 239			N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 470 470			N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 536 536			N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 607 607			N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 802 802			N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 899 899			N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 919 919			N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 965 965			N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 991 991			N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 1089 1089			N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 1196 1196			N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 1588 1588			N-LINKED (GLCNAC. .) (POTENTIAL).

P47037 saccharomyc  
P47460 mycoplasma  
Q9wva1 thermotoga  
P32380 saccharomyc  
P27895 saccharomyc  
P46468 plasmodium  
O58687 pyrococcus  
Q13201 homo sapien  
Q95293 dictyostelli  
P06101 saccharomyc  
P17119 saccharomyc  
P43573 saccharomyc

ALIGNMENTS

```

SQ SEQUENCE 1701 AA: 193719 MW: 3920B75E73D38552 CRC64;
Query Match 85.6%; Score 1946.5; DB 1; Length 1701;
Best Local Similarity 94.9%; Pred. No. 9.8e-93;
Matches 374; Conservative 7; Mismatches 8; Indels 5; Gaps 1;

QY 38 TDDDDKAMADIGSTEGRTMAISVTMDNILSGFENEYDVIYKPLAGYRSLKKQIEKNI 1347
DB 1293 SEDNDEYLDQVVTGE-----AISVTMDNILSGFENEYDVIYKPLAGYRSLKKQIEKNI 1347

QY 98 FTFNLDILNLSRLKRRKYFLDVLSDLMQFKHISSEYIIESDFKLLNSEQNTLLKS 157
DB 1348 ITFNLDILNLSRLKRRKYFLDVLSDLMQFKHISSEYIIESDFKLLNSEQNTLLKS 1407

QY 158 YKIKESVENDIKFAQEGISYEVKLVAKYKDDLESIKKVIKEEKEKFPSSPTTTPSPAK 217
DB 1408 YKIKESVENDIKFAQEGISYEVKLVAKYKDDLESIKKVIKEEKEKFPSSPTTTPSPAK 1467

QY 218 TDEQKESKFLPFLTNIETLYNNLVNKIDDDYLIINLKAKINDCNVEKDEAHVKITKLSDLK 277
DB 1468 TDEQKESKFLPFLTNIETLYNNLVNKIDDDYLIINLKAKINDCNVEKDEAHVKITKLSDLK 1527

QY 278 AIDDKIDLFKNPYDPEATKLLNDTKMDLKGKLLSTGLVONFNTIISKLECKFQDML 337
DB 1528 AIDDKIDLFKNPYDPEATKLLNDTKMDLKGKLLSTGLVONFNTIISKLECKFQDML 1587

QY 338 NISOHCVKKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGCCDADA 397
DB 1588 NISOHCVKKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGCCDADA 1647

QY 398 TCTEEDSGSSRRKKTCTCKTPDPSYPLDFGIFCSS 431
DB 1648 TCTEEDSGSSRRKKTCTCKTPDPSYPLDFGIFCSS 1681

RESULT 2
MSPL_PLAFM STANDARD; PRT; 1701 AA.
ID MSPL_PLAFM AC P08569;
DT 01-AUG-1988 (Rel. 08, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
DE (PMMSA) (P190).
GN MSP-1.
OS Plasmodium falciparum (isolate mad20 / Papua New Guinea).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=70153;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88011243; PubMed=3079521;
RA Tanabe K., Mackay M., Goman M., Scaife J.G.;
RT "Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmodium falciparum";
RT Plasmodium falciparum;
RL J. Mol. Biol. 195:273-287(1987).
RN [2]
RP REVISIONS TO 1403; 1569 AND 1629.
RA Tanabe K.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-115 FROM N.A.
RX MEDLINE=86136024; PubMed=3004972;
RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,
RA Stunnenberg H., Buyard H.;
RT "Polymorphism of the precursor for the major surface antigens of Plasmodium falciparum merozoites: studies at the genetic level.";
RL EMBO J. 4:3823-3829(1985).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (potential).
CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42 kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC -----
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
-----
EMBL: X05624; CAA29112.1; -
PIR: A26868; A26868.
PIR: B25120; B25120.
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein; Transmembrane; GPI-anchor.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1701 MEROZOITE SURFACE PROTEIN 1.
FT CARBOHYD 110 110 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 470 470 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 536 536 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 607 607 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 802 802 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 899 899 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 919 919 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 965 965 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 991 991 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1089 1089 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1196 1196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1588 1588 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1701 AA: 193768 MW: 3FC2EC59AF96EA98 CRC64;
Query Match 85.6%; Score 1946.5; DB 1; Length 1701;
Best Local Similarity 94.9%; Pred. No. 9.8e-93;
Matches 374; Conservative 7; Mismatches 8; Indels 5; Gaps 1;

QY 38 TDDDDKAMADIGSTEGRTMAISVTMDNILSGFENEYDVIYKPLAGYRSLKKQIEKNI 97
DB 1293 SEDNDEYLDQVVTGE-----AISVTMDNILSGFENEYDVIYKPLAGYRSLKKQIEKNI 1347

QY 98 FTFNLDILNLSRLKRRKYFLDVLSDLMQFKHISSEYIIESDFKLLNSEQNTLLKS 157
DB 1348 ITFNLDILNLSRLKRRKYFLDVLSDLMQFKHISSEYIIESDFKLLNSEQNTLLKS 1407

QY 158 YKIKESVENDIKFAQEGISYEVKLVAKYKDDLESIKKVIKEEKEKFPSSPTTTPSPAK 217
DB 1408 YKIKESVENDIKFAQEGISYEVKLVAKYKDDLESIKKVIKEEKEKFPSSPTTTPSPAK 1467

QY 218 TDEQKESKFLPFLTNIETLYNNLVNKIDDDYLIINLKAKINDCNVEKDEAHVKITKLSDLK 277
DB 1468 TDEQKESKFLPFLTNIETLYNNLVNKIDDDYLIINLKAKINDCNVEKDEAHVKITKLSDLK 1527

QY 278 AIDDKIDLFKNPYDPEATKLLNDTKMDLKGKLLSTGLVONFNTIISKLECKFQDML 337
DB 1528 AIDDKIDLFKNPYDPEATKLLNDTKMDLKGKLLSTGLVONFNTIISKLECKFQDML 1587

QY 338 NISOHCVKKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGCCDADA 397
DB 1588 NISOHCVKKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGCCDADA 1647

QY 398 TCTEEDSGSSRRKKTCTCKTPDPSYPLDFGIFCSS 431
DB 1648 TCTEEDSGSSRRKKTCTCKTPDPSYPLDFGIFCSS 1681
```



FT CARBOHYD 881 881 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 901 901 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 947 947 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1071 1071 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1178 1178 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1569 1569 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 1682 AA; 192462 MW; C82A1E159948CAD6 CRC64;

Query Match 85.1%; Score 1935; DB 1; Length 1682;  
 Best Local Similarity 94.4%; Pred. No. 3.8e-92;  
 Matches 372; Conservative 9; Mismatches 7; Indels 6; Gaps 2;

QY 38 TDDDDKAMADIGSTEGRTMAISVTMNLISGFENEVDVIYKPLAGVYRSKQKQIEKNI 97  
 Db 1275 SEDNEVDLQVVTGE-----AISVTMNLISGFENEVDVIYKPLAGVYRSKQKQIEKNI 1329  
 QY 98 FTFNLNLDIILSLKRRKRYFLDVLDSLMQFKHISNEYIIESFLLNSEQKNTLLKS 157  
 Db 1330 FTFNLNLDIILSLKRRKRYFLDVLDSLMQFKHISNEYIIESFLLNSEQKNTLLKS 1389  
 QY 158 YKIKESVENDIKFAQEGISYKVLAKYKDDLESIKKVIKEKEKFPSSPTTTPSPAK 217  
 Db 1390 YKIKESVENDIKFAQEGISYKVLAKYKDDLESIKKVIKEKEKFPSSPTTTPSPAK 1448  
 QY 218 TDEOKKESKFLPFTLTNIETLYNNLVNKIDDIYLNKAKINDCNVKEKDEAHVKITKLSDLK 277  
 Db 1449 TDEOKKESKFLPFTLTNIETLYNNLVNKIDDIYLNKAKINDCNVKEKDEAHVKITKLSDLK 1508  
 QY 278 AIDDKIDLFKNPYDFAIKKLINDTKKMLGKLLSTGLVQNFNTIISKLECKFQDML 337  
 Db 1509 AIDDKIDLFKNPYDFAIKKLINDTKKMLGKLLSTGLVQNFNTIISKLECKFQDML 1568  
 QY 338 NISQHCVKKQCPNSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGGCCDADA 397  
 Db 1569 NISQHCVKKQCPNSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGGCCDADA 1628  
 QY 398 TCTEEDSGSSRKKITCCTKRPDSYPLFDGIFCSS 431  
 Db 1629 KCTEEDSGSGKKITCCTKRPDSYPLFDGIFCSS 1662

RESULT 5  
 MSPL\_PLAAPP STANDARD; PRT; 1726 AA.  
 AC P50495;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Merozoite surface protein 1 precursor (Merozoite surface antigens) (PMSA) (GP195).  
 GN MSP-1.  
 OS Plasmodium falciparum (isolate Palo Alto / Uganda).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=57270;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89005525; PubMed=3049134;  
 RA Chang S.P., Kramer K.J., Yamaga K.M., Kato A., Case S.E.,  
 RA Siddiqui W.A.;  
 RT "Plasmodium falciparum: gene structure and hydrophathy profile of the  
 RT major merozoite surface antigen (gp195) of the Uganda-Palo Alto  
 RT isolate.";  
 RL Exp. Parasitol. 67:11-11(1988).  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor  
 CC (potential).  
 CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42  
 CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF  
 CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.  
 CC  
 CC This swiss-prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL: M37213; AAA29611.1; --  
 CC InterPro: IPR000561; EGF-like.  
 CC Pfam: PF00008; EGF; 1.  
 CC Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;  
 KW Transmembrane; GPI-anchor.  
 FT SIGNAL 1 19 POTENTIAL.  
 FT CHAIN 20 1726 MEROZOITE SURFACE PROTEIN 1.  
 FT CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 272 272 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 567 567 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 638 638 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 827 827 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 924 924 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 944 944 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 990 990 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1016 1016 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1114 1114 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1221 1221 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1613 1613 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 1726 AA; 196174 MW; 5B59CBEFA2F9A026 CRC64;

Query Match 84.9%; Score 1932.5; DB 1; Length 1726;  
 Best Local Similarity 94.2%; Pred. No. 5.2e-92;  
 Matches 371; Conservative 9; Mismatches 9; Indels 5; Gaps 1;

QY 38 TDDDDKAMADIGSTEGRTMAISVTMNLISGFENEVDVIYKPLAGVYRSKQKQIEKNI 97  
 Db 1318 SEDNEVDLQVVTGE-----AISVTMNLISGFENEVDVIYKPLAGVYRSKQKQIEKNI 1372  
 QY 98 FTFNLNLDIILSLKRRKRYFLDVLDSLMQFKHISNEYIIESFLLNSEQKNTLLKS 157  
 Db 1373 FTFNLNLDIILSLKRRKRYFLDVLDSLMQFKHISNEYIIESFLLNSEQKNTLLKS 1432  
 QY 158 YKIKESVENDIKFAQEGISYKVLAKYKDDLESIKKVIKEKEKFPSSPTTTPSPAK 217  
 Db 1433 YKIKESVENDIKFAQEGISYKVLAKYKDDLESIKKVIKEKEKFPSSPTTTPSPAK 1492  
 QY 218 TDEOKKESKFLPFTLTNIETLYNNLVNKIDDIYLNKAKINDCNVKEKDEAHVKITKLSDLK 277  
 Db 1493 TDEOKKESKFLPFTLTNIETLYNNLVNKIDDIYLNKAKINDCNVKEKDEAHVKITKLSDLK 1552  
 QY 278 AIDDKIDLFKNPYDFAIKKLINDTKKMLGKLLSTGLVQNFNTIISKLECKFQDML 337  
 Db 1553 AIDDKIDLFKNPYDFAIKKLINDTKKMLGKLLSTGLVQNFNTIISKLECKFQDML 1612  
 QY 338 NISQHCVKKQCPNSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGGCCDADA 397  
 Db 1613 NISQHCVKKQCPNSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGGCCDADA 1672  
 QY 398 TCTEEDSGSSRKKITCCTKRPDSYPLFDGIFCSS 431  
 Db 1673 KCTEEDSGSGKKITCCTKRPDSYPLFDGIFCSS 1706

RESULT 6  
 MSPL\_PLAAPP STANDARD; PRT; 1630 AA.  
 AC P04932;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Merozoite surface protein 1 precursor (Merozoite surface antigens) (PMSA) (P190).  
 GN MSP-1.  
 OS Plasmodium falciparum (isolate K1 / Thailand).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=5839;  
 RN [1]

SEQUENCE FROM N.A.  
 MEDLINE=86136024; PubMed=3004972;  
 RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,  
 RA Stunnenberg H., Bujard H.,  
 RT "Polymorphism of the precursor for the major surface antigens of  
 RT Plasmodium falciparum merozoites: studies at the genetic level.",  
 RL EMBO J. 4:3823-3829(1985).  
 RN [2]  
 RP REVISIONS, SEQUENCE FROM N.A.  
 RA Pan W., Tolle R., Bujard H.,  
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor  
 CC (potential).  
 CC PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42  
 CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF  
 CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X03371; CAA27070.1; -  
 DR PIR; A25120; SAZOK1.  
 DR InterPro: IPR000561; EGF-like.  
 DR Pfam: PF00008; EGF; 1  
 KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;  
 KW Transmembrane; GPI-anchor.  
 FT SIGNAL 1 19  
 FT CHAIN 20 1630  
 FT DOMAIN 67 84  
 FT TRANSMEM 1614 1630  
 FT CARBOHYD 97 97  
 FT CARBOHYD 259 259  
 FT CARBOHYD 755 755  
 FT CARBOHYD 759 759  
 FT CARBOHYD 774 774  
 FT CARBOHYD 835 835  
 FT CARBOHYD 911 911  
 FT CARBOHYD 955 955  
 FT CARBOHYD 1049 1049  
 FT CARBOHYD 1156 1156  
 FT CARBOHYD 1165 1165  
 FT CARBOHYD 1436 1436  
 FT CARBOHYD 1517 1517  
 FT SEQUENCE 1630 AA; 187289 MW; ADBEC3CE0A46322 CRC64;  
 Query Match 51.6%; Score 1175; DB 1; Length 1630;  
 Best Local Similarity 56.6%; Pred. No. 3.6e-53;  
 Matches 226; Conservative 59; Mismatches 90; Indels 24; Gaps 4;  
 QY 34 PDLGTDDDDKADIGSTEGRTMAISVTMDNILSGFENEYDVLYKPLAGVYRSLSKQI 93  
 DB 1235 PTFGESEED--YDDLGVQVVTGEAVTPSV-IDNILSKIENEYEVLYKPLAGVYRSLSKQI 1291  
 QY 94 EKNITFNLDNLNRLKRYKFLDVLDSLMOFKHHSSNEYIEDSFKLLNSEQKNT 153  
 DB 1292 ENNVTFNVNVDILNFRNKNENKFNVLSDLIIPKDTSSNVYVVKDPYKFLNKEKRD 1351  
 QY 154 LKSKYIKESVENDIKFAQEGISYERKVLAKYKDDLESIKKVIKEKEKFPSPPTTPP 213  
 DB 1352 FLSSYNYIKSDITDINFANDVLYGYYKILSEKYKSLDLSIKKYI----- 1395  
 QY 214 SPAKTDQKKSKFLPFTTETLTNNLNNLIDYLLNKKINDCNVEKDAHVYKYL 273  
 DB 1396 ----NDKQGENEYLPFLNNIETLYKTVDKIDLEVIHLEAKVLNVTYKSNVEVKIKEL 1451  
 QY 274 SDLKAIDKIDLFPKNPYDFEAIKLLINDTKKMLGKLLSTGLV-QNFPNTIISKLECK 332  
 DB 1452 NYLKTIQDKLADFKNNNVFVGIADLSTDYNNHNNLLTKFLSTGMVFENLAKTVLSNLLDGN 1511

QY 333 FODMLNISOHCVKQKOPENSQCFRHLDERECKCLLNKQEGDKCVENPPTCNENGG 392  
 DB 1512 LOGMLNISOHCVKQKOPENSQCFRHLDERECKCLLNKQEGDKCVENPPTCNENGG 1571  
 QY 393 CDADATCTEEDSGSSRRKITCTCKPDSYPLFDGIFCSS 431  
 DB 1572 CDADAKCTEEDSGSNGKKTCTCKPDSYPLFDGIFCSS 1610  
 RESULT 7  
 MSPL\_PLAFW  
 ID MSPL\_PLAFW STANDARD; PRT: 1639 AA.  
 AC P04933;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE Merozoite surface protein 1 precursor (Merozoite surface antigens)  
 DE (PMMSA) (P195).  
 GN MSP-1.  
 OS Plasmodium falciparum (isolate Wellcome).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OC NCBI\_TaxID=5848;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=86014355; PubMed=2995820;  
 RA Holder A.A., Lockyer M.J., Odink K.G., Sandhu J.S., Riveros-Moreno V.,  
 RA Nicholls S.C., Hillman Y., Davey L.S., Tizard M.L.V., Schwarz R.T.,  
 RA Freeman R.R.;  
 RT "Primary structure of the precursor to the three major surface  
 RT antigens of Plasmodium falciparum merozoites.",  
 RL Nature 317:270-273(1985).  
 RN [2]  
 RP REVISIONS.  
 RA Holder A.A.;  
 RL Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.  
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor  
 CC (potential).  
 CC PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42  
 CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF  
 CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X02919; CAA26676.1; -  
 DR PIR; A24594; A24594.  
 DR InterPro: IPR000561; EGF-like.  
 DR Pfam: PF00008; EGF; 1.  
 KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;  
 KW Transmembrane; GPI-anchor.  
 FT SIGNAL 1 19  
 FT CHAIN 20 1639  
 FT CARBOHYD 116 116  
 FT CARBOHYD 268 268  
 FT CARBOHYD 764 764  
 FT CARBOHYD 768 768  
 FT CARBOHYD 783 783  
 FT CARBOHYD 844 844  
 FT CARBOHYD 920 920  
 FT CARBOHYD 964 964  
 FT CARBOHYD 1058 1058  
 FT CARBOHYD 1165 1165  
 FT CARBOHYD 1174 1174  
 FT CARBOHYD 1445 1445  
 FT CARBOHYD 1526 1526  
 FT SEQUENCE 1639 AA; 187618 MW; 2C255B6616C87F6E CRC64;  
 SQ

```
Query Match      51.6%; Score 1175; DB 1; Length 1639;
Best Local Similarity 56.6%; Pred. No. 3.6e-53;
Matches 226; Conservative 59; Mismatches 90; Indels 24; Gaps 4;

QY 34 PDLGTDGDDKAMADIGSIEGRTWAIWTDNLSGSENEYDVLYLPLAGVYRSLAKQI 93
DB 1244 PIFGESEED--YDGLGVVTEAVTPSV-IDNLSKIENEYEVLYLPLAGVYRSLKQL 1300
QY 94 EKNFTFTNLNLDILNLSKRRKFLDVLSDLMQFHHISSNEYIIEDSKLLNSEOKNT 153
DB 1301 ENNVWTVNVVNDILNLSRFRKFNKFNKFNKFNKFNKFNKFNKFNKFNKFNK 1360
QY 154 LLKSYKIKESVENDIKFAQEGISYIEKVLAKYKDDLESIKKVKIKERKFPSPPTTP 213
DB 1361 FLSSYNIKDSIDYINFANDVLGYKILSEKYSOLDSIKKYI----- 1404
QY 214 SPKATDQKESKFLPFTNLTETLYNLSKIDYLLNLSKIDYLLNLSKIDYLLNLSK 273
DB 1405 ----NDKQGENEKLPLNLTETLYNLSKIDYLLNLSKIDYLLNLSKIDYLLNLSK 1460
QY 274 SLDKAIIDKIDLFKNPYDFAIKKLLINDTKKMLKLLSTGLV-QNFPNTIISKLEGG 332
DB 1461 NYLTIQDKLADFKNNFNKFNKFNKFNKFNKFNKFNKFNKFNKFNKFNKFNK 1520
QY 333 FQMLNISQHCQVKKQCPENSGCPRHLDERBECKLLNLYKQEGDKCVENPNTCNENGG 392
DB 1521 LQGLNISQHCQVKKQCPENSGCPRHLDERBECKLLNLYKQEGDKCVENPNTCNENGG 1580
QY 393 CDADATCTEEDSGSRKKTCTCKPDSYPLFDGIFCSS 431
DB 1581 CDADAKTEEDSGSRKKTCTCKPDSYPLFDGIFCSS 1619

RESULT 8
MSPL_PLAYO STANDARD; PRT; 1772 AA.
AC P13828;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
DE (PMMA) (230 kDa).
MSPL-1.
OS Plasmodium berghei yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5862;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90205979; PubMed=2320061;
RA Lewis A.P.;
RT "Sequence analysis upstream of the gene encoding the precursor to the
RT major merozoite surface antigens of Plasmodium yoelii.";
RL Mol. Biochem. Parasitol. 39:285-288(1990).
RN [2]
RP SEQUENCE OF 1093-1772 FROM N.A..
RC STRAIN=17XL;
RX MEDLINE=88124889; PubMed=2448778;
RA Burns J.M. Jr., Daly T.M., Vaidya A.B., Long C.A.;
RT "The 3' portion of the gene for a Plasmodium yoelii merozoite surface
RT antigen encodes the epitope recognized by a protective monoclonal
RT antibody.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:602-606(1988).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```



```
Db 461 -KLISQYESNFIKFNQNLHLSIFNLIQILQESSIAQFDRKMSIKSVNALKNLN 519
Qy 325 -----ISKLEKGFQDLNLSHQHCVKQKCPNSCFRHLDERECKLLNKKQBGD 376
Db 520 LIQPKLSLYFIETALESIINSYSLISMETPEP-----HOOGN 561
Qy 377 KCVENPN 383
Db 562 ELTATPN 568

RESULT 11
ID YL17_CABEL STANDARD; PRT: 1130 AA.
AC Q11102;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 131.5 kDa protein C02F12.7 in chromosome X.
GN C02F12.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Miller N.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: WEAK, TO MYOSINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U41545; AAK39135.1; .
DR Wormpep; C02F12.7; CE03901.
KW Hypothetical protein; Coiled coil.
FT DOMAIN 121 779 COILED COIL (POTENTIAL).
FT DOMAIN 805 1061 COILED COIL (POTENTIAL).
SQ SEQUENCE 1130 AA; 131485 MW; B0FD2EF3D99FB09 CRC64;

Query Match 6.4%; Score 145; DB 1; Length 1130;
Best Local Similarity 23.8%; Pred. No. 1.7;
Matches 78; Conservative 64; Mismatches 126; Indels 60; Gaps 15;

Qy 88 SLKQIEKNIPFNLDILNSRLK-----RYFLDVLSDLMQFKHLSN--EY 137
Db 596 NLRQLEKEI-SHTEDRNRLHENTQLEAHKETHETVRLAEAEIDQKSAFENEQY 654
Qy 138 IEDSFLLNSEQN-TLLSKYKIKESVENDIKFAQEGISYIEKVLAKYKDDLESIKV 196
Db 655 GKESAKIRELAQNKTLLSEMRVKHVAENLEAFTSDKNLLELESKN-KNIEHLKQE 713
Qy 197 IKEKEKFPSPPTTPSPAKTDQKESKFLPFTNLTLYNNLVNKIDYLYNLAKKI 256
Db 714 IAQLNEKI-----STKETKDELEKTIAQLEIDNSKSKQIEKHLRVNDML 761
Qy 257 NDCNVEKDEAHVTKTSLDKAIDDKID--LFKNPYDFEAIKKLINDDTKMDLGLKLLST 314
Db 762 DMGTIKDEL---VKNEEIKTISAKTQALLESNTVESETKLASVTEEREKE----- 810
Qy 315 GLVQNFNTIISKLECKFQDMLNISQ-----HOCVKKQCPNSGCFRHLDE-----REE 364
Db 811 --IQSP-OTQISE-----KNEVLTKAERINLETCLAKEREVELTGMRTKLDMDTQQLNEE 863
Qy 365 CKCLL--NYKQEGDKCVENPNPTCNENN 390
```

```
Db 864 TTVLVFDSNIOEK---IDKEKATINEMN 888

RESULT 12
ID YK26_YEAST STANDARD; PRT: 540 AA.
AC P36112;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Hypothetical 61.1 kDa protein in YP52-DBP7 intergenic region.
GN YKR016W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c;
RA Duesterhoeft A.; Moestl D.; Poehlmann R.; Philippsen P.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z28241; CAA82088.1; .
DR PIR; S38085; S38085.
DR SGD; S0001724; YKR016W.
KW Hypothetical protein.
SQ SEQUENCE 540 AA; 61082 MW; FD8715B29DB52507 CRC64;

Query Match 6.3%; Score 143; DB 1; Length 540;
Best Local Similarity 22.5%; Pred. No. 0.91;
Matches 103; Conservative 53; Mismatches 147; Indels 154; Gaps 22;

Qy 9 SGLVPRGSGMKETAATAKFEROHMDSPLDGTDD-----DDKAMADIGSIGEGTMAI 59
Db 68 SNNVPFAEDLLET-----YEHYH-DRPTLFLEDSWDGLKAKSNLDLLSGTSGQTSRNSRE 122
Qy 60 SVTMDNLS-----GFENEYDVIYKPLAGVYRSLKQIEKNIFTNLDILNSRLKRR 115
Db 123 NIEVKILSLPLENIETENS DPQLKEIIG-----SLNDILNSLNSN 164
Qy 116 KYFLDVLSDLMQFKHLSNRYEIIEDSFLLNSEQN-TLLSKYKIKESVENDIKFAQEG 175
Db 165 ---LSIPESFENSIKSNQN-----MLTNLSOLNETLK-----EA 196
Qy 176 ISYEVKVLAKYKDDLESIKKVIKEKEKFPSPPTTPSPAKTDQKESKFLPFTNIE 235
Db 197 LSNYM-----IQTSEVITELNTQYENS-----KREFE-----K 225
Qy 236 TLYNNLVNKIDYLYNL-----KAKINDCNVEKDEAH---VKITKLSDL-KAID.280
Db 226 NLQKNLLQEVDFENLTQKDKLEELKANEELQAKHANEVGLLSITQVKEFNKIK 285
Qy 281 DKIDLFRNP--YDFEAIKKLINDOTKK--DMLGKLLS--TGLVQ-----NF 320
Db 286 DKIEKERNGLRAHLEENINSEVNDLSKSIDRSKILSKNEALVQLTFQVDEIKSRINNLL 345
Qy 321 PNTIISKLECKFQDMLNISQHQVKKQCPNSGCFRHLDERECKCLLNKQEGD--- 376
Db 346 PDVWIDK--ELSLRLKLSNLLSTFNKSKCDDGDC-----CSCKKGNKNEGKEGI 394
Qy 377 --KCVENPNPTCNENNGCGDADATCTEEDSGSSRKKI 411
Db 395 SCCKPKPTNPSPILLSVALDELESTC-----SGKKI 424
```



```
RESULT 13
RA50_SULSO
ID RA50_SULSO STANDARD; PRT; 864 AA.
AC Q97WH0:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE DNA double-strand break repair rad50 ATPase.
GN RAD50 OR SSO2249.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2287;
RN [1].
SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11427726;
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Aweyaz M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RA "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC rad50/mre11 complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mre11 by unwinding
CC and/or repositioning DNA ends into the mre11 active site (By
CC similarity).
CC -!- SUBUNIT: Forms a complex with mre11 (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE006829; AAK42417.1;
CC InterPro; IPR003439; ABC transportr.
CC InterPro; IPR003405; SMC_C.
CC InterPro; IPR003395; SMC_N.
CC Pfam; PF02463; SMC_N; 1.
CC Pfam; PF02483; SMC_C; 1.
CC DNA repair; Hydrolyase; ATP-binding; Coiled coil; Complete proteome.
CC NP_BIND 30 37 ATP (BY SIMILARITY).
CC DOMAIN 164 701 COILED COIL (POTENTIAL).
CC SEQUENCE 864 AA; 101601 MW; 657076AE9B709FC CRC64;
Query Match 6.3%; Score 142.5; DB 1; Length 864;
Best Local Similarity 22.8%; Pred. No. 1.6;
Matches 87; Conservative 63; Mismatches 150; Indels 81; Gaps 16;
QY 27 ERQHMDSPDLGTDDDD-----KAMADIGSIEGRCGTMAISVTMDNLISGF 70
DB 364 ERLEKDAESELNDIDKVNLSLEQKVEPTRKKQLNLAQLAKE--SLISEKNEIINNISOV 421
QY 71 ENEYDIVYLKPLAGVYR-SLKQIEKNIFTNLDNLNLDLSRLKRRKYFLDVLESMDQF 129
DB 422 EGCTPCVGRPLDEEHKQIKIEAKSYILQLELNKNE-LEELKTKITNELNKIE---REY 477
QY 130 KHISNEYIIED---SFKLLNSEQKN--TLKSYKYIKESVENDIKFAQEGISYVEKV-- 182
DB 478 RRLSNKASYDNVMNRLQKLNNEIENLHSEIESLKNIDIEIKNEEVKELKLYEEFMR 537
QY 183 LAKY-KDDLESIK-----KVTKEEKEK-----FPSPPTTPPSAKTD 219
DB 538 LSKYTKKEELDKRVKLDENMKKKEETEKEMRGLESELKLDLRKALESKILDLNKRKVLKD 597
```

```
QY 220 EOKK-----ESKFLPFLTNI-----ETLYNNLVNKIDDYLINLRKAKIND 258
DB 598 EMKKKKGILEDYIRQVKLLQEEVKNLREEVNIQFDENRYNELKSLDAYNLSLKEKN- 656
QY 259 CNVEKDEAHVKTTL-SDLKAIDDKIDLFKNPYDFEAIKKLINDDTPKMKMLKLLSTGLV 317
DB 657 ---RKSRIEGELESLEKDIETISNRIANYE--LQLDKREKIINAINKLEKIRSAICERKL 711
QY 318 QNFPNPTIISKLEGGFQDMLN 338
DB 712 QSYIIMTTKQLIENNLNDIIS 732
```

```
RESULT 14
RA50_METJA
ID RA50_METJA STANDARD; PRT; 1005 AA.
AC Q58718;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE DNA double-strand break repair rad50 ATPase.
GN RAD50 OR MJ1322.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1].
SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=9637999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-P., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Furrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).
CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC rad50/mre11 complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mre11 by unwinding
CC and/or repositioning DNA ends into the mre11 active site (By
CC similarity).
CC -!- SUBUNIT: Forms a complex with mre11 (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U67572; AAB99331.1;
CC TIGR; MJ1322;
CC InterPro; IPR003439; ABC transportr.
CC InterPro; IPR003405; SMC_C.
CC InterPro; IPR003395; SMC_N.
CC Pfam; PF02463; SMC_N; 1.
CC Pfam; PF02483; SMC_C; 1.
CC ProDom; PD000006; ABC_transportr; 1.
CC DNA repair; Hydrolyase; ATP-binding; Coiled coil; Complete proteome.
CC NP_BIND 32 39 ATP (BY SIMILARITY).
CC DOMAIN 158 849 COILED COIL (POTENTIAL).
CC SEQUENCE 1005 AA; 119387 MW; 9BBB48173E788F3 CRC64;
Query Match 6.2%; Score 142; DB 1; Length 1005;
```



```

DR EMBL; AP000988; BAB67212.1; ALT_INIT.
DR InterPro; IPR003439; ABC_transportr.
DR InterPro; IPR003405; SMC_C.
DR InterPro; IPR003395; SMC_N.
DR Pfam; PF00005; ABC_tran; 1.
DR Pfam; PF02483; SMC_C; 1.
DR Pfam; PF02463; SMC_N; 1.
DR ProDom; PD000006; ABC_transportr; 1.
KW DNA repair; Hydrolase; ATP-binding; Coiled coil;
FT NP_BIND 30 37 ATP (BY SIMILARITY).
FT DOMAIN 172 731 COILED COIL (POTENTIAL).
SQ SEQUENCE 879 AA; 103487 MW; E9103B58914F20D7 CRC64;

Query Match      6.1%; Score 139.5; DB 1; Length 879;
Best Local Similarity 24.8%; Pred. No. 2.4; Mismatches 63; Gaps 14;
Matches 78; Conservative 55;

QY 88 SLKKQIEKNIFTNLDILNSLRKKRYFLDVLESLMQFKHSSNEYIIEEDSFKLLN 147
DB 250 SLNLKKDDISELREEVKDNRLREKEKLEKIDLEKD;---KLIEKEKEIIEAQNKIKL 305
QY 148 SEQKNTLLSKYKIVKESVENDIKFAQEGISYEKVLAKYKDDLESKKVIKEEKE-KFPS 206
DB 306 AOEKESKLTIKINLTDLSEKLRKRREEDYKKYI-EIKGELEL-----BEKERKENS 359
QY 207 SPPTTPPSPAKDEOKESKELPEFTNIETLYNNL-----VNKIDYLLNLKAKINDCNV 261
DB 360 -----LSDRKLSLKIK--LSEIESKISNRKISINIEELDELOKLNEDLNKKNQ 406
QY 262 EKDEAHVKITKLSLKAIDKIDLFKNPYDFFAIKKLINDDTKK-----DMLGKLLSTGL 316
DB 407 EREKL---ASQLGEIKG-----RIEELNKLGLNQVKNQVPCVGRSLDDH 451
QY 317 VQNFNPNTIISKL-----IEGKFQDMLN-----ISQ-HQCVKKQCPNSCFPHLDEREC 365
DB 452 KRKIQNEIEIKLELDELNKKFKLEINKTNGUISELNQIINKSKSEKDIATRNLADYNN- 510
QY 366 KCLLNYKQEGDKCQE 380
DB 511 --LLTQQQLRKEIE 523

RESULT 17
ID MYS2.DICD1 STANDARD; PRT; 2116 AA.
AC P08799;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin II heavy chain, non muscle.
GN MHCA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
OX NCBI_TaxId=44689;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87092266; PubMed=3540939;
RA Warrick H.M., de Lozanne A., Levinwand L.A., Spudich J.A.;
RT "Conserved protein domains in a myosin heavy chain gene from Dictyostelium discoideum.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:9433-9437(1986).
RC [2]
RP PHOSPHORYLATION SITES, AND MUTAGENESIS.
RC STRAIN=AX2;
RX MEDLINE=90353583; PubMed=2387408;
RA Lueck-Vielmeier D., Schleicher M., Grabatin B., Wippler J., Gerisch G.;
RT "Replacement of threonine residues by serine and alanine in a phosphorylatable heavy chain fragment of Dictyostelium myosin II.";
RL FEBS Lett. 269:239-243(1990).
RP [3]
PHOSPHORYLATION SITES.

```

```

RX MEDLINE=88112226; PubMed=28281113;
RA Wagle G., Noegel A., Scheel J., Gerisch G.;
RT "Phosphorylation of threonine residues on cloned fragments of the Dictyostelium myosin heavy chain.";
RL FEBS Lett. 227:71-75(1988).
RN [4]
RX X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 1-762.
RX MEDLINE=95345066; PubMed=7619795;
RA Fisher A.J., Smith C.A., Thoden J.B., Smith R., Sutoh K., Holden H.M., Rayment I.;
RT "X-ray structures of the myosin motor domain of Dictyostelium discoideum complexed with MgADP.Befx and MgADP.ALFA-";
RL Biochemistry 34:8960-8972(1995).
RN [5]
RX X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 1-762.
RX MEDLINE=95345067; PubMed=7619796;
RA Smith C.A., Rayment I.;
RT "X-ray structure of the magnesium(II)-pyrophosphate complex of the truncated head of Dictyostelium discoideum myosin to 2.7-A resolution.";
RL Biochemistry 34:8973-8981(1995).
RN [6]
RX X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
RX MEDLINE=96206189; PubMed=8611530;
RA Smith C.A., Rayment I.;
RT "X-ray structure of the magnesium(II).ADP.vanadate complex of the Dictyostelium discoideum myosin motor domain to 1.9-A resolution.";
RL Biochemistry 35:5404-5417(1996).
RN [7]
RX X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1-762.
RX MEDLINE=97452580; PubMed=9305951;
RA Gullick A.M., Bauer C.B., Thoden J.B., Rayment I.;
RT "X-ray structures of the MgADP, MgATPgammA, and MgAMPNP complexes of the Dictyostelium discoideum myosin motor domain.";
RL Biochemistry 36:11619-11628(1997).
RN [8]
RX X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
RX MEDLINE=98070605; PubMed=9405148;
RA Bauer C.B., Kuhlman P.A., Bageshaw C.R., Rayment I.;
RT "X-ray crystal structure and solution fluorescence characterization of Mg.2'(3')-O-(N-methylanthraniloyl) nucleotides bound to the Dictyostelium discoideum myosin motor domain.";
RL J. Mol. Biol. 274:394-407(1997).
CC -!- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN & HAS ATPASE ACTIVITY THAT IS ACTIVATED BY ACTIN.
CC -!- SUBUNIT: MYOSIN II HEAVY CHAIN IS TWO-HEADED. IT SELF-ASSEMBLES INTO FILAMENTS. HEXAMER OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -!- SUBCELLULAR LOCATION: HIGHEST CONCENTRATION IN THE POSTERIOR CELL CORTEX.
CC -!- DOMAIN: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HWM). IT CAN BE FURTHER SPLIT INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED SUBFRAGMENT (S2).
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES, CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -!- PTM: PHOSPHORYLATION INHIBITS THICK FILAMENT FORMATION AND REDUCES THE ACTIN-ACTIVATED ATPASE ACTIVITY.
CC -!- MISCELLANEOUS: DICTYOSTELIUM MYOSIN II HAS NO K(2)EDTA ATPASE ACTIVITY, PERHAPS CORRELATED WITH THE ABSENCE OF A CYS AT THE SH-1 POSITION (688).
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----

```

```

CC EMBL; M14628; AAA33227.1; -
DR PIR; A26655; A26655.
DR PIR; S00250; S00250.
DR PDB; 1MMA; 03-DEC-97.
DR PDB; 1MMD; 17-AUG-96.
DR PDB; 1MMG; 03-DEC-97.
DR PDB; 1MND; 03-DEC-97.
DR PDB; 1MNE; 17-AUG-96.
DR PDB; 1VOM; 17-AUG-96.
DR PDB; 1LVK; 28-JAN-98.
DR Dictyob; DD01008; mhca.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF00063; myosin_head.
DR Pfam; PF00612; IQ; 2.
DR Pfam; PF02736; Myosin_N; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR PRODOM; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS00096; IQ; 1.
DR Myosin; Coiled coil; Actin-binding; ATP-binding; 3D-structure;
KW Calmodulin-binding; Methylation; Alkylation; Phosphorylation.
FT DOMAIN 1 761
FT DOMAIN 762 791
FT DOMAIN 817 2116
FT NP_BIND 179 186
FT DOMAIN 638 660
FT DOMAIN 738 752
FT MOD_RES 130 130
FT MOD_RES 678 678
FT MOD_RES 1823 1823
FT MOD_RES 1833 1833
FT MOD_RES 2029 2029
FT SEQUENCE 2116 AA; 243871 MW; 2FC3770B81EE56A1 CRC64;

Query Match 6.1%; Score 138.5; DB 1; Length 2116;
Best Local Similarity 21.3%; Pred. No. 7.3;
Matches 75; Conservative

Qy 79 LKPLAGVYRSLKQIEKNIFTNLNLNLSRLKRRKYFLDV-LESLMQFKHISNEY 137
Db 1673 IKSLVAEYDEVEQLEDEI-----LAKDKLVKAKRALEVELEEVRLQLEEEEDSR 1723
Qy 138 IIEDSFLLNSEQNTLLSKYIKESVENDIKFAQEGISYKVLAKYKDDLESIKKVI 197
Db 1724 ELEDKRRLLTTEVED-ISKKI---DAVEQNTKL-----DEAKKLLTDDVTLKKQL 1771
Qy 198 KEEKEKFPSPPTTPPSPAKTDQKESKFLPFLTNIETLYNNLV-----NKIDDYLYN 251
Db 1772 EDEKKKLNES-----ERAKKRLESENEDFLAKDAEYKNSRAEKDRKKYKEDLKD 1822
Qy 252 LKAKINCNVKDEAHVKITKLSLDKIDDKLPNPYDEATKLLINDTKMDLGLK 311
Db 1823 TKYLNDEAATKTQTEIGAACL-----EDQIDELRSKLEQEQAKATQADKSKRTLEGEI 1876
Qy 312 LSTGLVNFNPNTIISKL-----TEGFQDMLNISQHCYKQKCPENSGCFRHL---DER 362
Db 1877 DNLRAQTEDEGKIWMLEKEKRALEGELEUEKTVTEAEADSKSAEOKSLRVLLELEDA 1936
Qy 363 EECKLLNYKQEGDKCVENPNPTNENNGGCDATCTEEDSGSRKKITCE 414
Db 1937 RNLQKEIDAKETAEADSKSNLQREIVEAKGRLE-EESARTNSDRSRKLEAE 1987

RESULT 18
SBCC_CLOAB
ID SBCC_CLOAB STANDARD; PRT; 1163 AA.
AC Q97FKL;
DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
GN Nuclease sbcd subunit C.
GN SBC OR CAC2736.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN-ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RA "Genome sequence and comparative analysis of the solvent-producing
FT bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838(2001).
CC -!- FUNCTION: Sbcd cleaves DNA hairpin structures. These structures
CC can inhibit DNA replication and are intermediates in certain DNA
CC recombination reactions. The complex acts as a 3'->5' double
CC strand exonuclease that can open hairpins. It also has a 5'
CC single-strand endonuclease activity (By similarity).
CC -!- SUBUNIT: Heterodimer of sbcc and sbcd (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. SBCC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE007771; AAK80682.1; -
DR InterPro; IPR003439; ABC_transportr.
KW Hydrolyase; Nuclease; Exonuclease; Endonuclease; DNA replication;
KW DNA recombination; ATP-binding; Coiled coil; Complete proteome.
FT NP_BIND 35 42
FT DOMAIN 197 415
FT DOMAIN 446 1003
FT SEQUENCE 1163 AA; 135507 MW; CESF0BD2215D7A92 CRC64;

Query Match 6.0%; Score 136; DB 1; Length 1163;
Best Local Similarity 19.8%; Pred. No. 5;
Matches 85; Conservative

Qy 64 DNI--LSGFENEYDVYIKPLAGVYRSLKQIEKNIFTNLNL-----NDI-----L 108
Db 677 DNIKDLSESNKIEVEYQKE-----KTVEKQCEKRIVDLKSELEAEIKERFNAVYTIENL 731
Qy 109 NSRLKKRYK---FDVLLESD-----LMQFKH-----ISS 134
Db 732 KAEKIQDFKEMKEILEKERVVEAEIGEIKDLNLRNLRIRTEKEQLMDKSRLEKLSK 791
Qy 135 NEYTIEDSFLLNS-----EQNTLLSKYIKESVENDIKFAQ-----174
Db 792 NKAELKEDKIINEKIELKKNKGVLDNLVELKEKIEGTIKKIEQYNLCDDKKMNEIEDK 851
Qy 175 -----GISYKVLAKYK-----DLESIKKVIKEEKKFSPPTTPPSPAKTDQKKE 224
Db 852 YRKCSDEIKYHNSLSLSDKRVNDIDKLNKILMEKPE-----NTEKAKENYLN 901
Qy 225 SKFLPFTNIETLYNNLVNKKIDDYLYNLRKAKINCNVKDE-----AHVKITKLSLDK 280
Db 902 DKEINLKSDEYKYNELSKNGAVEVLSKLLKNNRKLTEERKWIETQNNRVEKASKAKALQ 961
Qy 281 DKIDLFPNYPDEAIIKKLINDDTKMDLGLKLLSTGLVQNFNTI-----ISKLEEG-KFQ 334
Db 962 ER-----SIKLEEEVKNTIEIKELGKLLKT--KQELEHKLSSLDLDEKLFKGRKRV 1011
Qy 335 DMLNISQHCV-----KKQCPENSGCFRHLDERECKCLL-NYKQEGDKCVENPNPTCNE 389

```

```
Db 1012 EFVALNQLKYSTASKRLKEITGGNYGLEVDNDGKFIIRDYK----- 1054
QY 390 NGCCDADAT 398
Db 1055 NGAKRDAS 1063

RESULT 19
RBPL_PLAVB
ID RBPL_PLAVB STANDARD; PRT; 2869 AA.
AC Q00798;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Reticulocyte binding protein 1 precursor.
GN RBPL
OS Plasmodium vivax (strain Belem).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=31273;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92315338; PubMed=1617731;
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
merozoites."
RL Cell 69:1213-1226(1992).
CC -!- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
HUMAN RETICULOCYTE CELLS.
CC -!- SUBUNIT: HOMODIMER (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Membrane-bound.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL: M88097; AAA29743.1;
KW Malaria; Receptor; Signal; Transmembrane.
FT SIGNAL 1 17
FT CHAIN 18 2869 RETICULOCYTE BINDING PROTEIN 1.
FT DOMAIN 18 2807 EXTRACELLULAR.
FT TRANSMEM 2808 2826 POTENTIAL.
FT DOMAIN 2827 2869 CYTOPLASMIC.
FT SITE 1030 1032 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 2869 2869 AA; 330213 MW; 89DDBE442205EBCFF CRC64;
SQ SEQUENCE 2869 AA; 330213 MW; 89DDBE442205EBCFF CRC64;

Query Match 6.0%; Score 135.5; DB 1; Length 2869;
Best Local Similarity 21.9%; Pred. No. 15;
Matches 79; Conservative 70; Mismatches 136; Indels 75; Gaps 17;

QY 91 KQTEKNFTNLMNDI-----LNSRLKKRKYFLDVL-----ESDLMQFKHIS 133
Db 2265 RDIKKELYLFHQN-SDISIVEGGVQNMALYDLKNEKREMDLYRNISETKLQWEHST 2323
QY 134 SNEYIYEDSKLLNSQKNTLLKSYKIK-----ESVENDIKFAQGI-----SY 178
Db 2324 DVFKPMIELKHGNETNNKSLLEKEKLLKSVNDHMHSMEMKNGKLYTPESVQNNINI 2383
QY 179 YEKVLAKYKDDLESIKK-----VTKEEKFPSSPTTPSPAKTDEOKKESKFLPFL 231
Db 2384 YSVIEAEVK-TLEIDRDYGDNQIVIEHHKQFSILIDRTNALMDDIEFKKNNYLM 2442
QY 232 TNETLYNNLVNKIDYLLINKAKINDCNVKEDEAHVKITKLSLKAIDDKI-DLFKNPY 290
Db 2443 VNTET-----IHRVNDYIEKITNKLVOAKTEYEQI-----LENIKQNDMLQNIPLKV 2491
QY 291 D-----FPAIKK-----LINDTTKMDMLGKLLSTG-----LVONFPNTIISKLECKFDML 337
Db 291 D-----FPAIKK-----LINDTTKMDMLGKLLSTG-----LVONFPNTIISKLECKFDML 337
```

```
Db 2492 SIIEFYFNKKKESILNDLYEQE---RLKIGEHILDEIKRNVTEITSSYEIDQKMEMM- 2547
QY 338 NISOHOCVKKQCPENSGCFRHLDERECKLLANKQ--EGDKCVENPNTFCNENNGCDA 395
Db 2548 --SKNLEKSKMMNYTSIYEL-EREANEINRDAKQKDDTILNSVLEAAAIQKRGMDA 2604

RESULT 20
RA50_PYRAB
ID RA50_PYRAB STANDARD; PRT; 880 AA.
AC Q9UJC8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA double-strand break repair rad50 ATPase.
GN RAD50 OR PAR0812.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GE5 / Orsay;
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: Insights into archaeal chromosome
structure and evolution."
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
rad50/mre11 complex possesses single-strand endonuclease activity
and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mre11 by unwinding
and/or repositioning DNA ends into the mre11 active site (By
similarity).
CC -!- SUBUNIT: Forms a complex with mre11 (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL: AJ248286; CAB50131.1;
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003439; ABC_transportr.
DR InterPro: IPR003395; SMC_N.
DR Pfam: PF02463; SMC_N; 1.
DR ProDom: PD000006; ABC_transportr; 1.
DR SMART: SM00382; AAA; 1.
KW DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
FT NP_BIND 30 37 ATP (BY SIMILARITY).
FT DOMAIN 144 745 COILED COIL (POTENTIAL).
FT SEQUENCE 880 AA; 103970 MW; FDBI77EC7E026479 CRC64;

Query Match 5.9%; Score 135; DB 1; Length 880;
Best Local Similarity 24.0%; Pred. No. 4; 1;
Matches 79; Conservative 48; Mismatches 96; Indels 106; Gaps 13;

QY 79 LKPLAGYVRLSKKOIEKNFTNLMNDILNSRLKKRKYFLDVLSDLMQFKH----- 131
Db 237 LESIKGKISLKIQVEK-----LKGRK---KGLEEKIVQIERSEKKA 277
QY 132 -ISSNEVLIEDSKLLNSQKNTLLKSYKIKESVENDIKFAQGIISYKVLAKYKDDL 190
Db 278 KISELEIEIVDKIPKLOKEKEKRYKLKGFDEYES-----KLRLEKELSKWESEL 327
QY 191 ESTKKVKI-EKEKFPSSPTTPSPAKTDEOKKESKFLPFLTNIETLYNNL---VNKID 246
Db 328 KATEEVIKKEKKK-----ERAEIREK-----LSEIERLEELKAPYVELE 369
QY 247 DYLINKAKINDCNVKEDEAHVK-----ITKLSLKAIDDKIDLFPKNPYDFAIKKLI 299
```

Db 370 D-----AKQVQKQIERLKARLKGSPGVEIKLESLEKTERTEIE-----EAIKEI- 414  
 QY 300 NDDTKDMLGKLLTGLVQNPNTIISKLEGGKQDMLNLSHQVVKQCPENSCCPHLL 359  
 Db 415 -----TTRIGQMEQKERNKMAIELRAKAKCPV-----CGREL 449  
 QY 360 DERECKCLLNYKQGDGVNPNPTCNE 388  
 Db 450 TEHHKELMERYTLEIKKIEELKRTTEE 478

## RESULT 21

USO1\_YEAST STANDARD; PRT: 1790 AA.  
 AC P25386;  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 DE 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Intracellular protein transport protein USO1.  
 GN USO1 OR INT1 OR YDL058W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=X2180-1A;  
 RX MEDLINE=91185402; PubMed=2010462;  
 RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,  
 RA Yamasaki M.;  
 RT "A cytoskeleton-related gene, usol, is required for intracellular  
 RT protein transport in Saccharomyces cerevisiae.";  
 RL J. Cell Biol. 113:245-260(1991).  
 RN [2]  
 RP SEQUENCE OF 782-1790 FROM N.A.  
 RA Hostetter M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,  
 RA Kendrick K.E.;  
 RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 1-8 FROM N.A.  
 RA Bai Y., Symington L.S.;  
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI  
 CC COMPLEX.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR  
 CC MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE  
 CC ER AND THE GOLGI COMPLEX.  
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, COMPOSED  
 CC OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL  
 CC COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.  
 CC -1- SIMILARITY: BELONGS TO THE VDP/USO1/YBL047C FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X54378; CAA38253.1; -;  
 DR EMBL; L03188; AAB00143.1; -;  
 DR EMBL; U53668; AAB66659.1; -;  
 DR PIR; A38455; A38455.  
 DR SGD; S0002216; USO1.  
 DR InterPro; IPR002017; Spectrin.  
 KW Transport; Protein transport; Golgi stack; Cytoskeleton; Coiled coil.  
 FT DOMAIN 1 724 GLOBULAR HEAD.  
 FT DOMAIN 725 1790 COILED COIL (POTENTIAL).  
 FT DOMAIN 465 487 CHARGED (HYPER-HYDROPHILIC).  
 FT DOMAIN 991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.  
 FT DOMAIN 1172 1786 ASP/GLU-RICH (ACIDIC).

FT CONFLICT 847 847 G -> E (IN REF. 2).  
 FT CONFLICT 924 924 E -> K (IN REF. 2).  
 FT CONFLICT 1253 1253 V -> I (IN REF. 2).  
 FT CONFLICT 1319 1319 I -> V (IN REF. 2).  
 FT CONFLICT 1461 1461 N -> S (IN REF. 2).  
 FT CONFLICT 1581 1581 G -> S (IN REF. 2).  
 FT CONFLICT 1600 1600 I -> V (IN REF. 2).  
 FT CONFLICT 1661 1661 R -> S (IN REF. 2).  
 FT CONFLICT 1772 1772 D -> DEEDDEE (IN REF. 2).  
 SQ SEQUENCE 1790 AA; 206424 MW; 6CE2B216E9FD4818 CRC64;  
 Query Match 5.9%; Score 134; DB 1; Length 1790;  
 Best Local Similarity 21.2%; Pred. No. 10;  
 Matches 99; Conservative 77; Mismatches 191; Indels 100; Gaps 19;  
 QY 20 ETAAAKFERQHMDSPDLGTDGDDKAMADIGSIGEGRG-----TMAISVTMDNILSGFENE 73  
 Db 1165 EEQIANKEQYNEISQLNDEITSTQOENESIKKKKNDLEGEVAMKSTSEQSNLKKE 1224  
 QY 74 YDVIYLPAGVYRSLKQKQIEKNIFTNLDNLNLSRLKKRVFLDVLESDLMQFKHIS 133  
 Db 1225 IDALNLO-----IKELKKNETNEASLLESIKSVSEIVKIKEL-----QDECNFKKE 1273  
 QY 134 SNEYIIEFSKLNSEQKNT-----LLKSYKVIKESVE---NDIKPAQEGISYIEKVLAKY 186  
 Db 1274 VSE--LED--KLKASEDNKSKYLELQKESIKKEELDAKTTELKIQLEKITNLKAKESK 1329  
 QY 187 KDDLESIKKVIKEK-----EKFPSPPTTPSPAKTDQKKESKEL-PFLTNIELTY 238  
 Db 1330 ESELRLKKTSSERKNAEQLKLNKNE-----IQINKQAFERKRLNLSGGSTITQEQY 1383  
 QY 239 NNLVKNKIDDIYLI-----NLKAKINDCNVERDEARHKIT-----KLSDLKADDDKIDL 285  
 Db 1384 SEKINTLEDELIRLQENENELKAKEID-NTRSELEKVSLSNDELLEEKQNTIKSLQDEILS 1442  
 QY 286 FK-----NPDYFAIK-----KLINDTKKMDLGLKLL 312  
 Db 1443 YKDKITRNDKLLSTIERDNKRDLSEQLRAAESKAKVEEGLKLEESKKAELK 1502  
 QY 313 STGLVQNPNTIISKLEGGKQDMLNLSQH-----QCVKKQCPENSCGCPHLLDEBECKCL 368  
 Db 1503 SKEMMKLESTIESNETELK-SSMETIRKSEKLEQSKSAEEDIKNLOH-----EKSDDL 1556  
 QY 369 LNYKQEGDKCVENPNPCN-ENNGGCDADATCTEEDSGSSRRKKTCE 414  
 Db 1557 ISRINESEKQIEELKRLRIEAKSGSELETVKQELNNAQEKIRINAE 1603

## RESULT 22

MYS1\_YEAST STANDARD; PRT: 1928 AA.  
 AC P08964;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DE 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Myosin-I isoform (Type II myosin).  
 GN MYO1 OR YHR023W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C;  
 RX MEDLINE=91088308; PubMed=2263482;  
 RA Sweeney F.P., Watts F.Z., Pocklington M.J., Orr E.;  
 RT "The MYO1 gene from Saccharomyces cerevisiae: its complete nucleotide  
 RT sequence.";  
 RL Nucleic Acids Res. 18:7147-7147(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RX MEDLINE=94378003; PubMed=8091229;







CC and ATP-dependent double-strand-specific exonuclease activity.  
 CC Rad50 provides an ATP-dependent control of mre11 by unwinding  
 CC and/or repositioning DNA ends into the mre11 active site (By  
 CC similarity).  
 CC -!- SUBUNIT: Forms a complex with mre11 (By similarity).  
 CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.  
 CC  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AE007118; AAC07092.1; -  
 CC InterPro: IPR003439; ABC\_transport.  
 CC InterPro: IPR004592; SbcC.  
 CC InterPro: IPR002017; Spectrin.  
 CC TIGRFAMs: TIGR00618; sbcC.1.  
 KW DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.  
 FT NP\_BIND 32 39 ATP (BY SIMILARITY).  
 FT DOMAIN 160 826 COILED COIL (POTENTIAL).  
 SQ SEQUENCE 978 AA: 115897 MW: 980F2BF51ADD1151 CRC64:  
 Query Match 5.8%; Score 133; DB 1; Length 978;  
 Best Local Similarity 20.2%; Pred. No. 5.8;  
 Matches 94; Conservative 80; Mismatches 177; Indels 114; Gaps 21;  
 QY 10 GLVPR-----GSMKETAA-----KFRQHMSDPLGTDGDDKAMADI 48  
 Db 51 GKVPYRGASVATKYLVSERGEKALVDLDFSLGRNRYRVERFYREFP-----EDSQVRVYE- 105  
 QY 49 GSTEGRTWAISVTMDNI-----LSGFE-NEYDVIYLPAGVYSLRKKQIEKNFTFN 101  
 Db 106 ---EGR---RLNIKANEVEKWLKISGLDYKTFKTVILLPOGEFDFRLKESRKKILIN 159  
 QY 102 L-----NLNDILNRLKRRYFLDVLESDLMOFK-HISSNEYIIDSFKLLNSQKNT 153  
 Db 160 LLGLEELKVRQLASETFKNLECKREALKKEYELLDYTPTKREVLEKTLKNLEEE--- 215  
 QY 154 LKSYKIVIESVENDIKFAQEGISY---YEKVLAKYKDDLESIKKVIKEKEKFPSPPT 210  
 Db 216 -LKELTEKEUKLQKAEKDSLESLRSQVVKLK-LELENLEKEVEKLEKLEFSPKRV 273  
 QY 211 TTPSPAKTDEOKESKFLPFTNIETLYNNLVNK-----IDDYLNKAKINDCNVEDK--- 264  
 Db 274 APVPTAKRIEIDKK---LTELKVRKNKLTRELAVLDELSEFAQEELNRIEAEKEFK 329  
 QY 265 -----EARKVTIKLSLKAIDDKI-----DLFKNPYDFEATIKKILINDTKKMDLGLKL 312  
 Db 330 EEKEREKELEHLKLLQEIKEILKELSLSSLSLKEREREVEQAQEFDLSEVERKGGKL 389  
 QY 313 -----STGLVQNPNTIISKLEGKFD--MLNISQHCQVKQC 349  
 Db 390 VAEETKELEKIKELFSEETSLKMKRELVLQELQRLKELKEGEGLENLTQRYKEKKV 449  
 QY 350 PEN-----SCGFRHLDERE---ECKLLNVKQEGDKCVENPNPTC 386  
 Db 450 HKVNLNELLELERELRHLHYAHMVASVLSLPGDTC-----PVC 489  
 RESULT 25  
 ID MYS3\_SCHPO STANDARD; PRT: 2104 AA.  
 AC 014157; 042730;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Myosin type II heavy chain 2.  
 GN MYO3 OR MYP2 OR SPAC4A8.05C.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OX NCBI\_TaxID=4896;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98119390; PubMed=9459302;  
 RA Moteigi F., Nakano K., Kitayama C., Yamamoto M., Mabuchi I.;  
 RT "Identification of Myo3, a second type-II myosin heavy chain in the  
 RL fission yeast Schizosaccharomycetes pombe";  
 FEBS Lett. 420:161-166(1997).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98062346; PubMed=9398685;  
 RA Bezanilla M., Forsburg S.L., Pollard T.D.;  
 RT "Identification of a second myosin-II in Schizosaccharomycetes pombe:  
 RT Myo2p is conditionally required for cytokinesis";  
 Mol. Biol. Cell 8:2693-2705(1997).  
 RN (3)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeh J.,  
 RA Oliver K., O'Neill S., Pearson D., Quay L.A., Rabinowitsch E.,  
 RA Rutherford K., Rutter S., Saunders R., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volkart G., Aert R., Robben J., Grynoprez B.,  
 RA Weljens J., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel K., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
 RT "The genome sequence of Schizosaccharomycetes pombe";  
 Nature 415:871-880(2002).  
 RL Nature 415:871-880(2002).  
 CC -!- FUNCTION: STABILIZES THE F-ACTIN CABLES FORMING THE F-ACTIN RING  
 CC THAT SURROUNDS THE NUCLEUS DURING INTERPHASE. MAY WORK IN  
 CC CONJUNCTION WITH MYO2.  
 CC -!- SUBUNIT: BINDS TO CDC4 AND RLCL.  
 CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AB007633; BAA24579.1; -  
 DR EMBL: AF029788; AAC04615.1; -  
 DR EMBL: Z98762; CAB11475.1; -  
 DR HSP: P08799; 1MND.  
 DR InterPro: IPR004009; Myosin\_N.  
 DR InterPro: IPR001609; myosin\_head.  
 DR Pfam: PF00063; myosin\_head; 1.  
 DR Pfam: PF02736; Myosin\_N; 1.  
 DR PRINTS: PR00193; MYOSINHEAVY.  
 DR ProDom: PD000355; myosin\_head; 1.  
 DR SMART: SM00242; MYSC; 1.  
 KW Myosin; Coiled coil; ATP-binding; Actin-binding; Alkylation.

```
FT DOMAIN 1 828 MYOSIN HEAD-LIKE.
FT DOMAIN 829 2104 COILED COIL (POTENTIAL).
FT DOMAIN 646 660 ACTIN-BINDING (POTENTIAL).
FT NP_BIND 182 189 ATP (POTENTIAL).
FT MOD_RES 686 686 ALKYLATION (BY SIMILARITY).
FT CONFLICT 1193 1193 D -> G (IN REF. 2).
FT CONFLICT 1304 1304 E -> G (IN REF. 2).
FT CONFLICT 1344 1344 E -> K (IN REF. 2).
FT CONFLICT 1420 1420 G -> D (IN REF. 2).
SQ SEQUENCE 2104 AA; 242570 MW; 3A0548594028D258 CRC64;

Query Match 5.88; Score 133; DB 1; Length 2104;
Best Local Similarity 21.58; Pred. No. 14;
Matches 73; Conservative 55; Mismatches 122; Indels 90; Gaps 11;

QY 77 IYLPLAGVYRLSKKQTEKNIFTNLT-----NLNDILNRLKRRKYFLDVLDESLLMOPKH 131
Db 823 LHLAPLLGTTOTDEYLRKRDALINLNQLOESTKEVANELTITRERVLO-LTNDLQEEQA 881
QY 132 ISSNEYI-----IEDSKLLNSEQKNTLLSKYK-----IKESVENDIKFAQEGISY 178
Db 882 LAHEKDILVERANSRVEVHERLSLENOVTIADEKVEFLYAEOKOSTEEDLANKOTEISY 941
QY 179 YEKVLAKYKDLLESIKKVIKEKEKEFPSSPTTPPSAKTDEQKESKFLPFLNIETLY 238
Db 942 LSDLSSTLEKKLSIKK-----DEQTISSKY-----KELE 971
QY 239 NLNVNKIDDDY-----LINKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNPYDF 292
Db 972 KYLNIMADYQHSQHLNSLEKAINERNLNELNEKMLRL-----DDELLLKORSYDT 1025
QY 293 EAIKKLINDTKKMKGLKLLS--TGLVQNPNTTIISKLEGFQDMNLNISOHCVKKQCP 350
Db 1026 KQVELREENASLKQCRTYESQLASYSKYSET-----ESELNKEAEVLVIEQ-- 1073
QY 351 ENSGCFRHLDERECKLLNLYKQEGDKVENPNPTCENN 390
Db 1074 -----KEITEYRDQLHAKFQNPETHNIND 1098

RESULT 26
RA50_SULAC
ID RA50_SULAC STANDARD; PRT; 886 AA.
AC O33600;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA double-strand break repair rad50 Atpase.
GN RAD50
OS Sulfolobus acidocaldarius.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2285;
[1]
SEQUENCE FROM N.A.
RC STRAIN=ATCC 33909 / NCIB 11770 / DSM 639;
RA MEDLINE=97362314; PubMed=9211741;
RX Elie C., Baucher M.F., Fondrat C., Forterre P.;
RT "A protein related to eucaryal and bacterial DNA-motor proteins in the
hyperthermophilic archaeon Sulfolobus acidocaldarius.";
RL J. Mol. Evol. 45:107-114(1997).
CC -1- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
rad50/mrell complex possesses single-strand endonuclease activity
and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mrell by unwinding
and/or repositioning DNA ends into the mrell active site (By
similarity).
CC -1- SUBUNIT: Forms a complex with mrell (By similarity).
CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
```

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y10687; CAA71688.1; -
DR InterPro; IPR003439; ABC_transportr.
DR InterPro; IPR003395; SMC_N.
DR Pfam; PF02463; SMC_N; 1.
DR ProDom; PD000006; ABC_transportr; 1.
KW DNA repair; Hydrolase; ATP-binding; Coiled coil.
NP_BIND 30 37 ATP (BY SIMILARITY).
FT DOMAIN 174 727 COILED COIL (POTENTIAL).
SQ SEQUENCE 886 AA; 103857 MW; 0390AE1403194104 CRC64;

Query Match 5.88; Score 132.5; DB 1; Length 886;
Best Local Similarity 23.68; Pred. No. 5.5;
Matches 89; Conservative 67; Mismatches 142; Indels 79; Gaps 18;

QY 62 TMDNILSGFENEYDVIYLLKPLAGVYRSLKQIE-----KNITFNLNLDILN----- 109
Db 186 SIKDILKREAEID-----RLKKEIEIKVKLENIEREAKEKEDELNOYNTEF 233
QY 110 SRLKRRKYFLDVLSDLT-----MQPKHISSE-----YIIEDSKLL--NSEQKNT 153
Db 234 NRIKEIKVQVDILSGELSVNKKIEETALRLKDFEKEKRYNKIETEVKELDENREKINT 293
QY 154 L-----LKSQYIKYESVENDIKFAQEGISYVEKVLAKYK--DDLESIKKVIKEKE 202
Db 294 ISSFKSILVQIDSLKQINNVENDLKRKKEKLRKKELEKEKEQVEIEKRRKKELEKEK 353
QY 203 KFPSPPTTPPSAKTDEQKESKFLPFLNIETLYNNL-----VNKIDYLLNLKAKI 256
Db 354 QYEEIEKRLTYLVKNIERQKNEIEKLVV--DTQDLENKKIKDVSDRINQIDNELKGLDDR 412
QY 257 NDCNVEKDEAHVKITKLSDLKAI--DDKIDLFKNPYDEAIIKKLNDTCKDKMLKLLSTG 315
Db 413 GDLNGRKEQT-LKI--YNNLNSIEDRCPICGRPLDSEH--KAKIREETKVLLELNKQIT 468
QY 316 LVQNPNTTIISKLEGFQDMNLNISOHCVKKQCPENSGCF-----RHUDE-----RE 363
Db 469 ALQARINSLIKEREE--LEATRNKLQLELQKRS--KEGIIYEAKLKLQRLKEEKNKLQN 524
QY 364 ECKCLNLYKQEGDKCQE 380
Db 525 EILSLSYHQEFENIAE 541

RESULT 27.
RA50_PYRFU
ID RA50_PYRFU STANDARD; PRT; 882 AA.
AC P58301;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE DNA double-strand break repair rad50 Atpase.
GN RAD50 OR PF1167.
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
[1]
SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
RX MEDLINE=20485553; PubMed=11029422;
RA Hopfner K.-P., Karcher A., Shin D., Fairley C., Tainer J.A.,
Carney J.P.;
RT "Mrell and Rad50 from Pyrococcus furiosus: cloning and biochemical
characterization reveal an evolutionarily conserved multiprotein
machine.";
RL J. Bacteriol. 182:6036-6041(2000).
RN [2]
SEQUENCE FROM N.A.
```

RC STRAIN=Vcl1 / DSM 3638 / ATCC 43587 / JCM 8422;  
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;  
RT "The complete sequence of the *Pyrococcus furiosus* genome.";  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS) OF 1-149.  
RX STRAIN=Vcl1 / DSM 3638 / ATCC 43587 / JCM 8422;  
RX MEDLINE=2034838; PubMed=10892749;  
RA Hopfner K.-P., Karcher A., Shin D.S., Craig L., Arthur L.M.,  
RA Carney J.P., Tainer J.A.;  
RT "Structural biology of Rad50 ATPase: ATP-driven conformational  
RT control in DNA double-strand break repair and the ABC-ATPase  
RT superfamily.";  
RL Cell 101:789-800(2000).  
CC -1- FUNCTION: Involved in DNA double-strand break repair (DSBR). The  
CC rad50/mre11 complex possesses single-strand endonuclease activity  
CC and ATP-dependent double-strand-specific exonuclease activity.  
CC Rad50 provides an ATP-dependent control of mre11 by unwinding  
CC and/or repositioning DNA ends into the mre11 active site.  
CC -1- SUBUNIT: Forms a complex with mre11.  
CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AE010225; AAL81291.1; -  
CC PDB; 1F2T; 02-AUG-00.  
CC PDB; 1F2U; 02-AUG-00.  
CC InterPro: IPR003439; ABC transportr.  
CC InterPro: IPR003395; SMC\_N.  
CC Pfam: PF02463; SMC\_N; 1.  
KW DNA repair; Hydrolase; ATP-binding; Coiled coil; 3D-structure;  
FT Complete proteome.  
FT NP\_BIND 30 37 ATP.  
FT DOMAIN 148 744 COILED COIL (POTENTIAL).  
FT SQ SEQUENCE 882 AA; 103839 MW; 3ADCBD250382A99E CRC64;  
Query Match 5.8%; Score 132; DB 1; Length 882;  
Best Local Similarity 23.4%; Pred. No. 5.8;  
Matches 78; Conservative 73; Mismatches 101; Indels 82; Gaps 21;  
QY 52 EGRGTWAI5VTMDNLSGPNEDVLYLPLAGVYSLKKQIEKNFTFNLDILNSR 111  
Db 186 EYRDIILARTENIEELIK--ENEQELI-----OVLOEISK-----TEEVLPSS 225  
QY 112 LKRRYFLDVLSDLMQFHISSNEYIIEDSPKLLNSEQNTLLKSYKIKESVENDIKF 171  
Db 226 RSK-----VDMLREVLRL-----ETKVEIENSER---LLEKRGDKRLEGRKN 269  
QY 172 AQGISYIEKVLAKYKDDLESIKV--IKEKEKFPSPPTPPSPAKTDEQKSKFLP 229  
Db 270 TEE---YLEKLKEKELEBEQVEITSIKKDVAYLAL-----KEFKNEYLDKKYKIEK 320  
QY 230 FLTNIEFTLYNNLVNKIDY-----LINKAKIND--CNVEKD-EAHVKI-TKLSDL 276  
Db 321 ELTRVBEILNEIQKLEELNEKESEKEKLENEKLEIKLKAILEKHQHYEETKAKKENL 380  
QY 277 KAIIDKIDIFKNPYDFEAIKKLND-DTKKDMGLKLLSTGLVQNPFTIISKL--IEGKF 333  
Db 381 RQKELKLG-DKSPED---IKKLEELTKKT-----TIEERNEITQRIKELNKNI 427  
QY 334 QDMLN-ISOHQVKKQCPENSGFRHL--DEREE 364  
Db 428 GDLKTAIEELKAKGKCPV---CGRELTDEHREE 458  
RESULT 28  
BRC2\_MOUSE

102 BRC2\_MOUSE STANDARD; PRT; 3329 AA.  
AC P97929; P97383; O35922;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Breast cancer type 2 susceptibility protein.  
GN BRC2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=129;  
RX MEDLINE=97217789; PubMed=9063750;  
RA Connor F., Smith A., Wooster R., Stratton M., Dixon A., Campbell E.,  
RA Tait T.M., Freeman T., Ashworth A.;  
RT "Cloning, chromosomal mapping and expression pattern of the mouse  
RT Brc2 gene.";  
RL Hum. Mol. Genet. 6:291-300(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6;  
RX MEDLINE=97237041; PubMed=9119389;  
RA Sharan S.K., Bradley A.;  
RT "Murine Brc2: sequence, map position, and expression pattern.";  
RL Genomics 40:234-241(1997).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=129/SV;  
RX MEDLINE=97384941; PubMed=9242436;  
RA McAllister K.A., Haugen-Strano A., Hagevik S., Brownlee H.A.,  
RA Collins N.K., Futreal P.A., Bennett L.M., Wiseman R.W.;  
RT "Characterization of the rat and mouse homologues of the BRC2 breast  
RT cancer susceptibility gene.";  
RL Cancer Res. 57:3121-3125(1997).  
RN [4]  
RP SEQUENCE OF 18-200 FROM N.A.  
RX MEDLINE=97075121; PubMed=8917547;  
RA Rajan J.V., Wang M., Marquis S.T., Chodosh L.A.;  
RT "Brc2 is coordinately regulated with Brcal during proliferation and  
RT differentiation in mammary epithelial cells.";  
RL Proc. Natl. Acad. Sci. U.S.A. 93:13078-13083(1996).  
RN [5]  
RP SEQUENCE OF 569-625 FROM N.A.  
RX MEDLINE=97341126; PubMed=9196008;  
RA McAllister K.A., Ramchandran S., Haugen-Strano A., Fiedorek F.T. Jr.,  
RA Wiseman R.W.;  
RT "Genetic mapping of the Brc2 breast cancer susceptibility gene on  
RT mouse chromosome 5.";  
CC -1- FUNCTION: MAY PARTICIPATE IN A PATHWAY ASSOCIATED WITH THE  
CC ACTIVATION OF DOUBLE-STRAND BREAK REPAIR AND/OR HOMOLOGOUS  
CC RECOMBINATION (BY SIMILARITY).  
CC -1- SUBUNIT: INTERACTS WITH RAD51 (BY SIMILARITY).  
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED. HIGHEST EXPRESSION IN  
CC CEREBELLUM, TESTIS, ILEUM, APPENDIX, EPIDYMISS, OVARY AND MAMMARY  
CC GLAND. NO EXPRESSION IN LUNG.  
CC -1- DEVELOPMENTAL STAGE: IN THE MAMMARY GLAND, EXPRESSION INCREASES  
CC DRAMATICALLY DURING PREGNANCY.  
CC -1- SIMILARITY: CONTAINS 7 BRC2 REPEATS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U82270; AAB48306.1; -  
CC EMBL; U72947; AAB40720.1; -  
CC EMBL; U65594; AAC23702.1; -



```
Db 333 DOOSEYEHQHASFOYNRLKQLESSKANMDSIEYALN-----TKIVN----- 374
QY 312 LSTGLVQNPFTIISKL-TEGRFQDMLNISO 341
Db 375 -----LENRFESTWKEKNDIEEKYTALRSSE 401

RESULT 30
SMC1_YEAST
ID SMC1_YEAST STANDARD; PRT; 1225 AA.
AC P32908;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chromosome segregation protein SMC1 (DA-box protein SMC1).
GN SMC1 OR CHL10 OR YFL008W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94103320; PubMed=8276886;
RA Strunniukov A.V., Lartionov V.L., Koshland D.;
RT "SMC1: an essential yeast gene encoding a putative head-rod-tail
RT protein is required for nuclear division and defines a new ubiquitous
RT protein family.";
RL J. Cell Biol. 123:1635-1648(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RX MEDLINE=95400292; PubMed=7670463;
RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
RA Yamazaki M., Tashiro H., Eki T.;
RT "Analysis of the nucleotide sequence of chromosome VI from
RT Saccharomyces cerevisiae.";
RL Nat. Genet. 10:261-268(1995).
CC -1- FUNCTION: INVOLVED IN CHROMOSOME SEGREGATION IN MITOSIS. COULD BE
CC PART OF A CHROMOSOME CONDENSATION MOTOR.
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH SMC2 OR OLIGOMERS.
CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.
CC -1- DOMAIN: CONSISTS OF TWO PUTATIVE CENTRAL COILED-COIL REGIONS
CC FLANKED BY PUTATIVE GLOBULAR REGIONS AT THE N- AND C-TERMINUS.
CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L00602; AAA16595.1; -;
CC EMBL; D50617; BAA09230.1; -;
CC PIR; A49464; A49464.
CC PIR; S41804; S41804.
CC SGD; S0001886; SMC1.
CC InterPro; IPR003405; SMC_C.
CC InterPro; IPR003395; SMC_N.
CC Pfam; PF02463; SMC_N; 1.
CC Pfam; PF02483; SMC_C; 1.
CC MitoSis; ATP-binding; Coiled coil; Nuclear protein.
CC NP_BIND 33 40
CC FT DOMAIN 173 489
CC FT COILED COIL (POTENTIAL).
CC FT DOMAIN 679 1063
CC FT COILED COIL (POTENTIAL).
CC FT DOMAIN 1057 1061
CC FT NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC FT DOMAIN 1137 1164
CC FT ALA/ASP-RICH (DA-BOX).
CC FT MUTAGEN 173 173
CC FT S->L: IN TS MUTANT SMC1-2.
CC FT MUTAGEN 458 458
CC FT N->D: IN TS MUTANT SMC1-1.
CC SEQUENCE 1225 AA; 141279 MW; B504017AA0ECCA8C CRC64;
```

```
Query Match 5.8%; Score 131.5; DB 1; Length 1225;
Best Local Similarity 20.9%; Pred. No. 9;
Matches 93; Conservative 72; Mismatches 126; Indels 155; Gaps 23;

QY 38 TDDDDKAMADIGSIEGR-----GTMAISVTMDNILSGFENEYDVII--LKPLAGVY 86
Db 253 TDKLSALNSEISLKGKINNMKSLQSKSFVKESAVISKQSKLDYIFDKREKLVSDL 312
QY 87 RSLK-----KQTEKNFTFNLDNLNRLKRYKRYFLDVLSDLMQFKHISNE 136
Db 313 RLKVPQQAAGKRKISHIEKRI-----ESLQKDLDQRTYVERFET---QLKVVTRSK 361
QY 137 YIIE-----DSFKLLNSEQNFKLLSKYIKESVENDIKFAQEGISYVEKVLAKY 186
Db 362 EAFEEEIKQSARNYDKFKLNEND-----LKYNCLEH-----KYLTEGGSILKEIAVL 410
QY 187 KDLESIKKKVKEKEKFPSPPTTPPSPAKTDEQKESKFLPLTNIETLYNNLVNKID 246
Db 411 NND-----KREIQEELERFNK-----RADISKRRITEELSGITG-----EKLD 447
QY 247 DYLLINKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKN-PYD----- 291
Db 448 TQLNDLRVSLN-----EKNALHTE---RUHLKLLQSDIESANNQYDNLFKLRETLVKIDD 501
QY 292 -----FEAKKLIND--DTKKDMLGKLLSTGLVQNPFTII 325
Db 502 LSAHQRETMKERKLRENIAMLKRFFPGVGLVHDLCHPKKRYGLAVSTILGKNFDSVIV 561
QY 326 SKLIEGKFQDMLNISQHCV---KKQCPENSGGCFRHLDERE-ECKCL----- 368
Db 562 -----ENLTVAQ-ECTAFLLKKQ-RAGTASFIPDLTIETELPTLSLPDSQDYIILSIN 610
QY 369 -LNVKQEGDKCQE---NPNPTCNENN 390
Db 611 AIDYEPYERAMQYVCGDSIICTNLN 636

RESULT 31
RBP2_PLAVB
ID RBP2_PLAVB STANDARD; PRT; 1251 AA.
AC Q00759;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Reticulocyte binding protein 2 (Fragment).
GN RBP2.
OS Plasmodium vivax (strain Belem).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=31273;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=92315338; PubMed=1617731;
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
RT merozoites.";
RL Cell 69:1213-1226(1992).
CC -1- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
CC HUMAN RETICULOCYTE CELLS.
CC -1- SUBCELLULAR LOCATION: Membrane-bound (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M88098; AAA29744.1; -;
CC KW Malaria; Receptor; Membrane.
CC FT NON_TER 1251 1251
CC FT SEQUENCE 1251 AA; 143741 MW; 54BA51C7404AC572 CRC64;
```



RA Peterson S., Reich C.I., McNeill L.K., Badger J.H., Glodek A., Zhou L.,  
 RA Overbeek R., Cocayne J.D., Weidman J.F., McDonald L., Utterback T.,  
 RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,  
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,  
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
 RA Venter J.C.;  
 RT "The complete genome sequence of the hyperthermophilic, sulphate-  
 RT reducing archaeon *Archaeoglobus fulgidus*.";  
 RL Nature 390:364-370(1997).  
 CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The  
 CC rad50/mre11 complex possesses single-strand endonuclease activity  
 CC and ATP-dependent double-strand-specific exonuclease activity.  
 CC Rad50 provides an ATP-dependent control of mre11 by unwinding  
 CC and/or repositioning DNA ends into the mre11 active site (By  
 CC similarity).  
 CC -!- SUBUNIT: Forms a complex with mre11 (By similarity).  
 CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE001032; AAB90211.1; -;  
 DR TIGR; AF1032; -;  
 DR InterPro; IPR003439; ABC\_transportr.  
 DR InterPro; IPR003395; SMC\_N.  
 DR InterPro; IPR02017; Spectrin.  
 DR Pfam; PF02463; SMC\_N; 1.  
 DR DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.  
 FT NP\_BIND 31 38  
 FT DOMAIN 148 728  
 FT SEQUENCE 886 AA; 103633 MW; D35641D499A8B58 CRC64;  
 SQ  
 Query Match 5.7%; Score 130; DB:1; Length 886;  
 Best Local Similarity 21.6%; Pred. No. 7.4;  
 Matches 85; Conservative 63; Mismatches 128; Indels 118; Gaps 18;  
 QY 47 DIGSI-----EGRGTMAISVTMDNILSGFENEYDVYVLPKPLAGVYRSLKKQIEKNITFNL 102  
 Db 130 EIDSIIRDDSERIIRIQTR---IEDYENAM-----KNLGAVIRMLEREKER----- 174  
 QY 103 NLDILN--SRLLKRYFLDVLESMLQFKHSSNEYIYEDSFKLNSQKN----- 152  
 Db 175 -LKEFLSQEQLKQK-----EKKAEIERISEEKSIESLREKLSSEVRNLESRLKEL 227  
 QY 153 ----TLKSYKYIKESVENDIKFAQBGISYEVKVLAKYKDDLESIKKVIKEKEKPPSP 208  
 Db 228 EEHKSRLSEKQESSVLQEVGRGLEKLEKLEKQKVEVRIEDLEKAKVEKLPKAE 287  
 QY 209 P-----TTPSPAKTDQKE-----SKFLPFLNIETLYNN 240  
 Db 288 RVSILEKLLSEINQALRDVEKREGDLTREAAAGTQALQKAEEDNSLEITRIEELERE 347  
 QY 241 L-----VNKIDYLNILNKAINDCNVKEKDEAHVKITKLSDL--KAIDDKIDLF 286  
 Db 348 LERFEKSHRLLETLPKPDPMOGIKAKLEKNLTPD-----KVEKMYDLLSKAKEEKEKI- 402  
 QY 287 KNPYDFEAIKLLINDTDPKMDMLKGLSTGLVQNFPPNTIISKLTIEGKQFMDLNISQHCVK 346  
 Db 403 -----TEKLKLLI---AKSS-----LKTGAQ-----LKKAVE-----ELKSAE 434  
 QY 347 KQCPENSGCFRLHDEREECKLLNLYKQEGDKCIVE 380  
 Db 435 RTCPV---CGRELDEEHRKNMAEYTRMKRTAE 465  
 RESULT 34  
 SMC3\_YEAST  
 ID SMC3\_YEAST STANDARD; PRT; 1230 AA.

AC P47037;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Chromosome segregation protein SMC3 (DA-box protein SMC3).  
 GN SMC3 OR YJL074C OR J1049.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=WB303;  
 RX MEDLINE=97474309; PubMed=9335333;  
 RA Michaelis C., Ciosek R., Nasmyth K.;  
 RT "Cohesins: chromosomal proteins that prevent premature separation of  
 RT sister chromatids.";  
 RL Cell 91:35-45(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC Rose M., Koetter P., Entian K.D.;  
 RA Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288c;  
 RA Sor F.J.;  
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: INVOLVED IN CHROMOSOME SEGREGATION IN MITOSIS. COULD BE  
 CC PART OF A CHROMOSOME CONDENSATION MOTOR.  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -!- DOMAIN: CONSISTS OF TWO PUTATIVE CENTRAL COILED-COIL REGIONS  
 CC FLANKED BY PUTATIVE GLOBULAR REGIONS AT THE N- AND C-TERMINUS.  
 CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Y14278; CAA74655.1; -;  
 DR EMBL; Z49349; CAA89366.1; -;  
 DR EMBL; X88851; CAA61313.1; -;  
 DR SGD; S0003610; SMC3.  
 DR InterPro; IPR003439; ABC\_transportr.  
 DR InterPro; IPR003405; SMC\_C.  
 DR InterPro; IPR003395; SMC\_N.  
 DR Pfam; PF02463; SMC\_N; 1.  
 DR Pfam; PF02483; SMC\_C; 1.  
 KW Mitosis; ATP-binding; Coiled coil; Nuclear protein.  
 FT NP\_BIND 32 39  
 FT DOMAIN 172 482  
 FT DOMAIN 685 1041  
 FT SEQUENCE 1230 AA; 141336 MW; B152D88F780341F CRC64;  
 SQ  
 Query Match 5.7%; Score 130; DB:1; Length 1230;  
 Best Local Similarity 19.6%; Pred. No. 11;  
 Matches 92; Conservative 83; Mismatches 169; Indels 126; Gaps 17;  
 QY 11 LVPRGCMKETAATAAKFERQHMDSPDLGTD-DKAMADIGSIEGRGTMAISVT-----MD 64  
 Db 145 IVPQGVKVALTNKDKERQLLEDVVGAKEFVKLAKSKKMEETQKKIQINKEMGELN 204  
 QY 65 NILSGFENEYDVYVLPKPLAGVYRSLKKQIEK-----NIFTNL---NLNIDLSRLK 113  
 Db 205 SKLSEMEQE-----RKELEKYNELERNRKIYQFTLYDRELNEVINOMER 248  
 QY 114 KRKYFLDVLESMLQFKHSSNEYIYEDSFKLNSQKNITLKKSYKIKE----- 163  
 Db 249 LDGDYNTVYSSEQYTQELDKREDMDIQVSKLSSIEASLUKIKNATDLOQAKLRESEISQ 308

```
QY 164 ---SVENDIKFAQGISYEVKVLAKYKDDLESIKKVIKEKEKPPSPPTTPPSAKTDE 220
DB 309 KLTWNVNIKVOQOIESNEQORNLSATLKEIKSIIEQRKOKL-----SKILPRYOEL 362
QY 221 QKESKELPFLTNIETLYNNLVNKIDDY-----LINLKAINDCNVEKD 264
DB 363 TKEEAMVKLQASQOQKORDILKKGAYARKSKDERTWTHSEIEELKSSQNLNELES 422
QY 265 EAHVITKL-SDLKAIDDKIDLFKNPYDFAEIKKLINDTKKMLGKLLSTGLVQNPNT 323
DB 423 QLQMDRTSLRKQYSAIDEIE-----ELIDSGPDTK-----GOLEDDESE 464
QY 324 II---SKLIEG-----KFDML-----NISOHQCVKQCPENSCFPHLDE-- 361
DB 465 LIHLKQKLSLSDTRKELWRKEQKLOTVLETLFLSDVANOQ-----RNVNETM 511
QY 362 -REECKLLNKBQEGDKVENPNP---TCNENNGCDAATCTEEDSGS 407
DB 512 SRSLANGLIINVKTEKLIKSPESVFTGLGELIKVNDKYKTCAEVIGNS 561

RESULT 35
HMW2_MYCGE STANDARD; PRT; 1805 AA.
AC P47460;
DT 01-FEB-1996 (Rel. 33, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Cytadherence high molecular weight protein 2 (Cytadherence accessory
DE protein 2)
GN HMW2 OR MG218.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 33530 / G-37;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.F., Dougherty B.A., Bitt K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RA "The minimal gene complement of Mycoplasma genitalium.";
RA Science 270:397-403(1995).
RN [2]
RP SEQUENCE OF 557-659 FROM N.A.
RC STRAIN-ATCC 33530 / G-37;
RA MEDLINE-96026346; PubMed-7569993;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RA "Evidence for lateral gene transfer between Archaea and Bacteria from
RA genome sequence of Thermotoga maritima.";
RA Nature 399:323-329(1999).
CC -!- FUNCTION: NOT KNOWN.
CC -!- SIMILARITY: BELONGS TO THE GIDA FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U39701; AAC71437.1;
CC EMBL; U02165; AAD12447.1;
CC TIGR; MG218;
CC
CC Cytadherence; Structural protein; Coiled coil; Complete proteome.
```

```
FT DOMAIN 28 838 COILED COIL (POTENTIAL).
FT DOMAIN 914 1591 COILED COIL (POTENTIAL).
FT DOMAIN 1632 1723 COILED COIL (POTENTIAL).
FT DOMAIN 1777 1804 COILED COIL (POTENTIAL).
SQ SEQUENCE 1805 AA; 216252 MW; 11D093AF173284FD CRC64;

Query Match 5.7%; Score 130; DB 1; Length 1805;
Best Local Similarity 21.2%; Pred. No. 17;
Matches 84; Conservative 66; Mismatches 124; Indels 122; Gaps 20;

QY 25 KFERQHMDSPDLGTDODDKANADIGSTEGRTMAISVTMDNILSGFENEYD--VIYLLKPL 82
DB 1191 KFEKONFD-----TEKQQLVAIKTQCEKLSDEKKALNOKLVELKNL 1232
QY 83 AGVYRSLLKQIE-----KNIFTNLNL---NDILNSRLKRRKYF--LDVLESDL- 126
DB 1233 SQTILANKNAEYSSQOQLOQKYTNLLDLKENLERTKQDLK--KRSIFARLTFFANDLR 1290
QY 127 MQFHSSNEYIIEDSPKLNSEQKNTLLSKYKIKESVENDIKFAQGISYVEK----- 181
DB 1291 FEKKOLLKAQRIVDKKNRLLKENERN-----LHFLSNETERKRAVLEDOISYFEKQKQA 1345
QY 182 ---VLAKYKO-----DLESIKKVIKEKEKFPSPPTTPPSAKTDEOKKE----- 224
DB 1346 TDAILASHKEVKKKEGELQKLLVELTRKTKLNN-----DFAKFSQREEFENQRLKL 1398
QY 225 ---SKFLPFTNIETLYNNLVNKIDDDYL-----INLKAINDCN-----V 261
DB 1399 LELOKTLQOTNSNFKTKAQIEIENSYKRGMEELNQKKEFDKNSRLYEFPRKMRDEI 1458
QY 262 EKDEAHVKIT-----KLSDLKAIDDKIDLFKNPYDF-----EAIKKLIN---DQTKK-- 305
DB 1459 ERKESQVKLVLETKQRKANLLEAQANKLINIEKNTIDFEKELKAFKQKVDQDIDSTNQR 1518
QY 306 ---DML--GKLLSTGLVQ-----NFPNTIISKLIE 330
DB 1519 KELNELLNENKLLQOOLIERERAINSKDSLUNKKIE 1554

RESULT 36
GIDA_THEMA STANDARD; PRT; 629 AA.
AC O9KYAL;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Glucose inhibited division protein A.
GN GIDA OR TM0263.
OS Thermotoga maritima
OC Bacteria; Thermotogae; Thermotogae (class); Thermotogales;
OC Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MSE8 / DSM 3109;
RX MEDLINE-99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RA "Evidence for lateral gene transfer between Archaea and Bacteria from
RA genome sequence of Thermotoga maritima.";
RA Nature 399:323-329(1999).
CC -!- FUNCTION: NOT KNOWN.
CC -!- SIMILARITY: BELONGS TO THE GIDA FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U39701; AAC71437.1;
CC EMBL; U02165; AAD12447.1;
CC TIGR; MG218;
CC
CC Cytadherence; Structural protein; Coiled coil; Complete proteome.
```





Db 616 EISLSKSI-DRYKKDF-----NQLKSE--QSNIQHDLNLQILNLNKL-- 656

QY 351 ENSGCFRHLDERECKLLNYKQEGDKVCNPNPTCNENGGCDADATCTEEDSGSSRKK 410

Db 657 -----IESEDELSLRDSQK-----IETENKRYNNLSLENDRLLTKEASDKER 703

RESULT 38

CIN8\_YEAST

ID CIN8\_YEAST STANDARD; PRT; 1038 AA.

AC P27895;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE Kinesin-like protein CIN8.

GN CIN8 OR KSL2 OR YEL061C.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI\_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=S288C;

RX MEDLINE=92317149; PubMed=1618897;

RA Hoyt M.A., He L., Loo K.K., Saunders W.S.;

RT "Two Saccharomyces cerevisiae kinesin-related gene products required

RT for mitotic spindle assembly.";

RL J. Cell Biol. 118:109-120(1992).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=S288C / AB972;

RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,

RA Aviles E., Berno A., Brennan T., Carpenter J., Chen E., Cherry J.M.,

RA Chung E., Duncan M., Guzman E., Hartzell G., Hunnicke-Smith S.,

RA Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,

RA Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,

RA Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,

RA Taylor P., Wei Y., Yelton M., Bolstein D., Davis R.W.;

RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE OF 1021-1038 FROM N.A.

RC STRAIN=S288C;

RX MEDLINE=9517238; PubMed=7867803;

RA Rousselet G., Simon M., Ripoch P., Buhler J.M.;

RT "A second nitrogen permease regulator in Saccharomyces cerevisiae.";

RL FEBS Lett. 359:215-219(1995).

RN [4]

RP CHARACTERIZATION.

RC STRAIN=S288C;

RX MEDLINE=92354062; PubMed=1643659;

RA Saunders W.S., Hoyt M.A.;

RT "Kinesin-related proteins required for structural integrity of the

RT mitotic spindle.";

RL Cell 70:451-458(1992).

CC -1- FUNCTION: REQUIRED FOR ASSEMBLY OF THE MITOTIC SPINDLE. INTERACT

CC WITH SPINDLE MICROTUBULES TO PRODUCE AN OUTWARDLY DIRECTED

CC FORCE ACTING UPON THE POLES. FOLLOWING SPINDLE ASSEMBLY, CIN8 AND

CC KIP1 APPARENTLY ACT TO OPPOSE A FORCE THAT DRAWS SEPARATED POLES

CC BACK TOGETHER. THIS FORCE SEEMS TO BE MEDIATED BY KAR3.

CC -1- SUBCELLULAR LOCATION: SPINDLE MICROTUBULES THAT LIE BETWEEN THE

CC POLES.

CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. BINC

CC SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DB EMBL; Z11859; CAA77885.1; -

DR EMBL; M90522; AAA34496.1; -

DR EMBL; U18795; AAB65026.1; -

DR EMBL; X79105; CAA55722.1; -

DR PIR; B42641; B42641.

DR HSP; P17119; 3KAR.

DR SGD; S0000787; CIN8.

DR InterPro; IPR001752; kinesin\_motor.

DR Pfam; PF00225; kinesin; 1.

DR PRINTS; PR00380; KINESINHEAVY.

DR SMART; SM00129; KISC; 1.

DR PROSITE; PS00411; KINESIN\_MOTOR\_DOMAIN1; 1.

DR PROSITE; PS00667; KINESIN\_MOTOR\_DOMAIN2; 1.

KW Motor protein; Cell division; Microtubules; ATP-binding; Coiled coil;

KW Mitosis; Cell cycle.

FT DOMAIN 72 553 KINESIN-MOTOR (BY SIMILARITY).

FT DOMAIN 554 657 COILED COIL (POTENTIAL).

FT DOMAIN 904 942 COILED COIL (POTENTIAL).

FT NP\_BIND 166 173 ATP (BY SIMILARITY).

FT CONFLICT 254 254 D -> A (IN REF. 1).

FT CONFLICT 831 831 Q -> H (IN REF. 1).

SQ SEQUENCE 1038 AA; 117999 MW; 3A1FD7003EF89FBC CRC64;

Query Match 5.6%; Score 128; DB 1; Length 1038;

Best Local Similarity 24.3%; Pred. No. 11;

Matches 89; Conservative 54; Mismatches 109; Indels 114; Gaps 21;

QY 86 YRSLLKQIEKNIPTFNLDILNSRLKRRKYFLDVLSDLMQFHHISSNE--YIIDSF 143

DB 508 YAKAKNIKPKQLGSGFTMKDIL---VKNITMELAKIKSDLLSTK---SKEGIYMSODHY 561

QY 144 KLLNSEQNTLLSKYKIKESVENDIKFAQEGISYEVKVLAKYKDDLESIKVKKEEK 203

DB 562 KNLNSD-----LESYKNEVQECKREIE-----SLTSKNALLVKDKLKS-KETIQSQNCQ 609

QY 204 FPSPPPTTPSPAKTDEOKESKFLPFLTNTE-TLYNNLVNKIDY----LINKAKIND 258

DB 610 IESLTKTTIDHLRAQLDKQHK-----TEISDFNNKLOKLTVEVQMALHDYKKRELD 661

QY 259 CNVEKDEAHVKITK-LSDLKA-----IDDKIDLFKNP----- 289

DB 662 LN-QKFEH--ITREIKKLKSTFLQLNTMOQESILQETNIQPNLDMIKNEVLTLMTMQ 718

QY 290 -----YDFAIKKLINDTKDMLGLLSTGLVQNFNTIISK--LIEGFO----- 334

DB 719 EKALMYK-DCVKILNESPK-----FFNVVIEKIDIIIRVDFQKFKYKIA 762

QY 335 -DMLNISQHCQVKKOCPENSGCF-----RHLDERECKCLLYKQEGDKVCNPN 383

DB 763 ENLSDISEENNMMKOYLKNIH--PFKNNHQELLNRHVDSTYE-----NIEKRTNEFVFNPK 815

QY 384 PTCNEN 389

DB 816 KVLNDH 821

RESULT 39

CDAT\_PLAFA STANDARD; PRT; 709 AA.

ID CDAT\_PLAFA

AC P46468;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 01-NOV-1995 (Rel. 32, Last annotation update)

DE Putative cell division cycle ATPase (Fragment).

OS Plasmodium falciparum.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.

OX NCBI\_TaxID=5933;

RN [1]

RP SEQUENCE FROM N.A.

RA Sims P.F., Hyde J.E.;

RL Submitted (XXX-1993) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration





GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM:protein - protein search, using sw model

Run on: May 12, 2003, 10:15:19 ; Search time 88 Seconds  
(without alignments)  
1009.163 Million cell updates/sec

Title: US-10-057-531A-2

Perfect score: 2275

Sequence: 1 MHHHHHHSSGLVPRGSMKE.....TCECTKPDVPLFDGIFCSS 431

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_21.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phase.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_rvirus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1969.5	86.6	1720	5 Q25922	Q25922 plasmodium
2	1965.5	86.4	652	5 Q25923	Q25923 plasmodium
3	1946.5	85.6	570	5 Q9TYG2	Q9TYG2 plasmodium
4	1946.5	85.6	570	5 Q25968	Q25968 plasmodium
5	1942	85.4	373	5 Q25723	Q25723 plasmodium
6	1937	85.1	569	5 Q25978	Q25978 plasmodium
7	1935	85.1	373	5 Q25722	Q25722 plasmodium
8	1933	85.0	373	5 Q25727	Q25727 plasmodium
9	1931	84.9	569	5 Q25970	Q25970 plasmodium
10	1931	84.9	569	5 Q25980	Q25980 plasmodium
11	1931	84.9	569	5 Q25982	Q25982 plasmodium
12	1929	84.8	373	5 Q43996	Q43996 plasmodium
13	1928	84.7	373	5 Q25728	Q25728 plasmodium
14	1924	84.6	373	5 Q25724	Q25724 plasmodium
15	1922.5	84.5	372	5 Q25717	Q25717 plasmodium
16	1921	84.4	569	5 Q25983	Q25983 plasmodium

17	1919	84.4	373	5	Q43995	plasmodium
18	1919	84.4	373	5	Q25721	plasmodium
19	1917.5	84.3	372	5	Q25718	plasmodium
20	1917.5	84.3	372	5	Q25719	plasmodium
21	1917.5	84.3	372	5	Q25720	plasmodium
22	1917	84.3	651	5	Q25924	plasmodium
23	1915	84.2	569	5	Q25969	plasmodium
24	1915	84.2	569	5	Q25974	plasmodium
25	1915	84.2	569	5	Q25975	plasmodium
26	1915	84.2	569	5	Q25977	plasmodium
27	1915	84.2	569	5	Q25979	plasmodium
28	1915	84.2	1694	5	Q9NHX1	plasmodium
29	1915	84.2	1694	5	Q9TGT5	plasmodium
30	1915	84.2	1704	5	Q9TGT4	plasmodium
31	1906.5	83.8	372	5	Q25725	plasmodium
32	1901.5	83.6	372	5	Q25726	plasmodium
33	1900.5	83.5	372	5	Q43997	plasmodium
34	1900	83.5	569	5	Q25967	plasmodium
35	1178	51.8	539	5	Q25972	plasmodium
36	1178	51.8	539	5	Q25981	plasmodium
37	1175	51.6	400	5	Q03999	plasmodium
38	1175	51.6	539	5	Q25984	plasmodium
39	1175	51.6	539	5	Q25966	plasmodium
40	1175	51.6	539	5	Q9TYG1	plasmodium
41	1175	51.6	539	5	Q25976	plasmodium
42	1169	51.4	539	5	Q25971	plasmodium
43	1169	51.4	539	5	Q25973	plasmodium
44	1165.5	51.2	376	5	Q9BMG8	plasmodium
45	1049	46.1	218	5	Q9TVG8	plasmodium

#### ALIGNMENTS

#### RESULT 1

Q25922 ID Q25922 PRELIMINARY; PRT; 1720-AA.  
AC Q25922;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)  
DE Precursor of the major merozoite surface antigens.  
OS Plasmodium falciparum (isolate NF54).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5843;  
RN [1]  
RP SEQUENCE OF 1069-1720 FROM N.A.  
RC STRAIN=NF54;  
RA Tolle R., Bujard H., Cooper J.A.;  
RL Exp. Parasitol. 0:0-0(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NF54;  
RA Tolle R.;  
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NF54;  
RX MEDLINE=96123395; PubMed=8577332;  
RA Pan W., Tolle R., Bujard H.;  
RT "A direct and rapid sequencing strategy for the Plasmodium falciparum antigen gene gp190/MSA1";  
RL Mol. Biochem. Parasitol. 73:241-244(1995).  
DR EMBL: 235327; CA84556.1;  
DR InterPro: IPR000561; EGF-like.  
DR Pfam: PF00008; EGF; 1.  
KW Merozoite; Signal.  
FT SIGNAL 1 19 POTENTIAL.  
SQ SEQUENCE 1720 AA; 195725 MW; 717B2FD1E637A8A3 CRC64;

Query Match 86.6%; Score 1969.5; DB 5; Length 1720;  
Best Local Similarity 95.7%; Pred. No. 8.7e-88;  
Matches 377; Conservative 7; Mismatches 5; Indels 5; Gaps 1;

```
QY 38 TDDDKAMADIGSIEGRGTMAISVTMDNILSGFENEYDVYILKPLAGYVRSLLKQKQIEKNI 97
Db 1312 SEDNDYLDQVVTGE-----AISVTMDNILSGFENEYDVYILKPLAGYVRSLLKQKQIEKNI 1366
QY 98 FTFNLNDILNSRLKRRKRYFLDVLSDLMQFKHISSEYIIEDSFKLNSEQKNTLLKS 157
Db 1367 FTFNLNDILNSRLKRRKRYFLDVLSDLMQFKHISSEYIIEDSFKLNSEQKNTLLKS 1426
QY 158 YKIKESVENDIKFAQEGISYKVLAKYKDDLESIKKVIKEEKEKPPSPPTTPPSPAK 217
Db 1427 YKIKESVENDIKFAQEGISYKVLAKYKDDLESIKKVIKEEKEKPPSPPTTPPSPAK 1486
QY 218 TDEQKESKFLPFLTNIETLYNNLVNKIDDYLYNLKAKINDCNVEKDEAHVKITKLSDLK 277
Db 1487 TDEQKESKFLPFLTNIETLYNNLVNKIDDYLYNLKAKINDCNVEKDEAHVKITKLSDLK 1546
QY 278 AIDDKIDLFKNPYDFAEIKKLLINDTKKMLGKLLSTGLVQNFNTIISKLEIGKFDQML 337
Db 1547 AIDDKIDLFKNPYDFAEIKKLLINDTKKMLGKLLSTGLVQNFNTIISKLEIGKFDQML 1606
QY 338 NISOHCVKKQCPNSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENNGCCDADA 397
Db 1607 NISOHCVKKQCPNSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENNGCCDADA 1666
QY 398 TCTEEDSGSRKKITCCTKPDSPYPLDFGIFCSS 431'
Db 1667 TCTEEDSGSRKKITCCTKPDSPYPLDFGIFCSS 1700
```

## RESULT 2

```
Q25923 ID Q25923 PRELIMINARY; PRT; 652 AA.
AC Q25923;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE Merozoite surface antigen 1 (Fragment).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE OF 1-298 FROM N.A.
RC STRAIN=FCH5/NF7;
RA MEDLINE=88142999; PubMed=2449612;
RA Peterson M.G., Coppel R.L., McIntyre P., Langford C.J., Woodrow G.,
RA Brown G.V., Anders R.F., Kemp D.J.;
RT "Variation in the precursor to the major merozoite surface antigens of
RT Plasmodium falciparum.";
RL J. Mol. Biochem. Parasitol. 27:291-302(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FCH5/NF7;
RA Tolle R.;
RT Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=FCH5/NF7;
RA Tolle R., Bujard H., Cooper J.A.;
RT "Plasmodium falciparum: variations within the C-terminal region of
RT merozoite surface antigen-1.";
RL EMBL; 235328; CAA84557.1;
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
FT NON_TER 1
SQ SEQUENCE 652 AA; 74292 MW; 2B6A87737B490A62 CRC64;
```

Query Match 86.4%; Score 1965.5; DB 5; Length 652;  
Best Local Similarity 95.4%; Pred. No. 4.9e-88;  
Matches 376; Conservative 8; Mismatches 5; Indels 5; Gaps 1;

```
QY 38 TDDDKAMADIGSIEGRGTMAISVTMDNILSGFENEYDVYILKPLAGYVRSLLKQKQIEKNI 97
Db 244 SEDNDYLDQVVTGE-----AISVTMDNILSGFENEYDVYILKPLAGYVRSLLKQKQIEKNI 298
QY 98 FTFNLNDILNSRLKRRKRYFLDVLSDLMQFKHISSEYIIEDSFKLNSEQKNTLLKS 157
Db 299 FTFNLNDILNSRLKRRKRYFLDVLSDLMQFKHISSEYIIEDSFKLNSEQKNTLLKS 358
QY 158 YKIKESVENDIKFAQEGISYKVLAKYKDDLESIKKVIKEEKEKPPSPPTTPPSPAK 217
Db 359 YKIKESVENDIKFAQEGISYKVLAKYKDDLESIKKVIKEEKEKPPSPPTTPPSPAK 418
QY 218 TDEQKESKFLPFLTNIETLYNNLVNKIDDYLYNLKAKINDCNVEKDEAHVKITKLSDLK 277
Db 419 TDEQKESKFLPFLTNIETLYNNLVNKIDDYLYNLKAKINDCNVEKDEAHVKITKLSDLK 478
QY 278 AIDDKIDLFKNPYDFAEIKKLLINDTKKMLGKLLSTGLVQNFNTIISKLEIGKFDQML 337
Db 479 AIDDKIDLFKNPYDFAEIKKLLINDTKKMLGKLLSTGLVQNFNTIISKLEIGKFDQML 538
QY 338 NISOHCVKKQCPNSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENNGCCDADA 397
Db 539 NISOHCVKKQCPNSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENNGCCDADA 598
QY 398 TCTEEDSGSRKKITCCTKPDSPYPLDFGIFCSS 431
Db 599 TCTEEDSGSRKKITCCTKPDSPYPLDFGIFCSS 632
```

## RESULT 3

```
Q9TYG2 ID Q9TYG2 PRELIMINARY; PRT; 570 AA.
AC Q9TYG2;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88011243; PubMed=3079521;
RA Tanabe K., Mackay M., Goman M., Scalfe J.G.;
RT "Allelic Dimorphism in a Surface Antigen Gene of the Malaria Parasite
RT Plasmodium falciparum.";
RL J. Mol. Biol. 195:273-287(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13343; BAA02604.1;
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 570 AA; 64630 MW; 8674DEC89B2D662A CRC64;
```

Query Match 85.6%; Score 1946.5; DB 5; Length 570;  
Best Local Similarity 94.9%; Pred. No. 3.6e-87;  
Matches 374; Conservative 7; Mismatches 8; Indels 5; Gaps 1;

```
QY 38 TDDDKAMADIGSIEGRGTMAISVTMDNILSGFENEYDVYILKPLAGYVRSLLKQKQIEKNI 97
Db 162 SEDNDYLDQVVTGE-----AISVTMDNILSGFENEYDVYILKPLAGYVRSLLKQKQIEKNI 216
QY 98 FTFNLNDILNSRLKRRKRYFLDVLSDLMQFKHISSEYIIEDSFKLNSEQKNTLLKS 157
```

```
Db 217 ITFNLNLNDILNSRLKKRYFLDVLESDLMOFKHHSNEYIEDSFKLNSQKNTLLKS 276
Qy 158 YKIKESVENDIKFAQEGISYYEKVLAKYKDDLESIKKVIKEKEKFPSSPPTTPSPAK 217
Db 277 YKIKESVENDIKFAQEGISYYEKVLAKYKDDLESIKKVIKEKEKFPSSPPTTPSPAK 336
Qy 218 TDEQKESKFLPFLNITETLYNNLVNKIDDDYLINLAKINDCNVDEKDEAHVKITKLSDLK 277
Db 337 TDEQKESKFLPFLNITETLYNNLVNKIDDDYLINLAKINDCNVDEKDEAHVKITKLSDLK 396
Qy 278 AIDDKIDLFPKPYDEALTKKLLNDTCKDMLGKLLSTGLVQNFPTIISKLEGGKFDML 337
Db 397 AIDDKIDLFPKPYDEALTKKLLNDTCKDMLGKLLSTGLVQNFPTIISKLEGGKFDML 456
Qy 338 NISQHCYKVKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGGCDADA 397
Db 457 NISQHCYKVKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGGCDADA 516
Qy 398 TCTEDSGSSRRKKTCTCKTDPSPFLPDGIFCSS 431
Db 517 TCTEDSGSSRRKKTCTCKTDPSPFLPDGIFCSS 550

RESULT 4
Q25968
ID Q25968 PRELIMINARY; PRT; 570 AA.
AC Q25968;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93293445; PubMed=8515786;
RA Jongwutiwes S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates";
RL MOL. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13345; BAA02606.1; -
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 570 AA; 64632 MW; 424BF553CCC2F2BE CRC64;

Query Match 85.6%; Score 1946.5; DB 5; Length 570;
Best Local Similarity 94.9%; Pred. No. 3.6e-87;
Matches 374; Conservative 7; Mismatches 8; Indels 5; Gaps 1;

Qy 38 TDDDDKAMADIGSIEGRGTMAISVTMDNLTSFGFENEYDVIYKPLAGVYRSLLKQIEKNI 97
Db 162 SDNDDEYLDQVVGTG-----AISVTMDNLTSFGFENEYDVIYKPLAGVYRSLLKQIEKNI 216
Qy 98 FTFNLNLDILNSRLKKRYFLDVLESDLMOFKHHSNEYIEDSFKLNSQKNTLLKS 157
Db 217 ITFNLNLNDILNSRLKKRYFLDVLESDLMOFKHHSNEYIEDSFKLNSQKNTLLKS 276
Qy 158 YKIKESVENDIKFAQEGISYYEKVLAKYKDDLESIKKVIKEKEKFPSSPPTTPSPAK 217
Db 277 YKIKESVENDIKFAQEGISYYEKVLAKYKDDLESIKKVIKEKEKFPSSPPTTPSPAK 336
Qy 218 TDEQKESKFLPFLNITETLYNNLVNKIDDDYLINLAKINDCNVDEKDEAHVKITKLSDLK 277
Db 337 TDEQKESKFLPFLNITETLYNNLVNKIDDDYLINLAKINDCNVDEKDEAHVKITKLSDLK 396
Qy 278 AIDDKIDLFPKPYDEALTKKLLNDTCKDMLGKLLSTGLVQNFPTIISKLEGGKFDML 337
Db 397 AIDDKIDLFPKPYDEALTKKLLNDTCKDMLGKLLSTGLVQNFPTIISKLEGGKFDML 456
Qy 338 NISQHCYKVKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGGCDADA 397
Db 457 NISQHCYKVKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGGCDADA 516
Qy 398 TCTEDSGSSRRKKTCTCKTDPSPFLPDGIFCSS 431
Db 517 TCTEDSGSSRRKKTCTCKTDPSPFLPDGIFCSS 550
```

```
Qy 338 NISQHCYKVKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGGCDADA 397
Db 457 NISQHCYKVKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGGCDADA 516
Qy 398 TCTEDSGSSRRKKTCTCKTDPSPFLPDGIFCSS 431
Db 517 TCTEDSGSSRRKKTCTCKTDPSPFLPDGIFCSS 550

RESULT 5
Q25723
ID Q25723 PRELIMINARY; PRT; 373 AA.
AC Q25723;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1, 42 kDa C-terminal region (Fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=INFECTED HUMAN BLOOD SAMPLE;
RA Shi Y.-P., Alpers M.P., Povea M.M., Nahlen B.L., Oloo A.G., Lal A.A.;
RT "Sequence of the C-terminal region of merozoite surface protein-1
RT (MSP-1) in field-derived Plasmodium falciparum";
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U20728; AAG62219.1; -
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
FT NON_TER 1
SQ SEQUENCE 373 AA; 42848 MW; EE9A891631DE174F CRC64;

Query Match 85.4%; Score 1942; DB 5; Length 373;
Best Local Similarity 98.9%; Pred. No. 3.8e-87;
Matches 369; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 58 AISVTMDNLTSFGFENEYDVIYKPLAGVYRSLLKQIEKNI FTFNLNLDILNSRLKKRY 117
Db 1 AISVTMDNLTSFGFENEYDVIYKPLAGVYRSLLKQIEKNI FTFNLNLDILNSRLKKRY 60
Qy 118 FLDVLESDLMOFKHHSNEYIEDSFKLNSQKNTLLKSYKIKESVENDIKFAQEGIS 177
Db 61 FLDVLESDLMOFKHHSNEYIEDSFKLNSQKNTLLKSYKIKESVENDIKFAQEGIS 120
Qy 178 YYEKVLAKYKDDLESIKKVIKEKEKFPSPPTTPSPAKTDEQKESKFLPFLNITETL 237
Db 121 YYEKVLAKYKDDLESIKKVIKEKEKFPSPPTTPSPAKTDEQKESKFLPFLNITETL 180
Qy 238 YNNLVNKIDDDYLINLAKINDCNVDEKDEAHVKITKLSDLKAIDDKIDLFPKPYDFAIKK 297
Db 181 YNNLVNKIDDDYLINLAKINDCNVDEKDEAHVKITKLSDLKAIDDKIDLFPKPYDFAIKK 240
Qy 298 LINDDTCKDMLGKLLSTGLVQNFPTIISKLEGGKFDMLNISOHCYKVKQCPENSGCFR 357
Db 241 LINDDTCKDMLGKLLSTGLVQNFPTIISKLEGGKFDMLNISOHCYKVKQCPENSGCFR 300
Qy 358 HLDERECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRRKKTCTCKT 417
Db 301 HLDERECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSGKKITCTCKT 360
Qy 418 PDSYPLFDGIFCS 430
Db 361 PDSYPLFDGIFCS 373

RESULT 6
Q25978
ID Q25978 PRELIMINARY; PRT; 569 AA.
AC Q25978;
```





```
Db 1 AISVTMDNLISGFENEYDVIIYKPLAGVYRSLLKKQIEKNIITFNLDNLNSRLKRRY 60
QY 118 FLDVLESOLMQPKHSSNEYYIETDSFKLNSQKNTLLSKYKIKESVENDIKFAOEGIS 177
Db 61 FLDVLESOLMQPKHSSNEYYIETDSFKLNSQKNTLLSKYKIKESVENDIKFAOEGIS 120
QY 178 YEKVLAKYKDDLESIKKVIKEKEKFPSPPTTPSPAKTDEQKSKFPLPLTNIETL 237
Db 121 YEKVLAKYKDDLESIKKVIKEKEKFPSPPTTPSPAKTDEQKSKFPLPLTNIETL 180
QY 238 YNNLVNKIDDYLLNLKAKINDCNVDEAHVKITKLSLDLKAIDDKIDLKFNPDFAIKK 297
Db 181 YNNLVNKIDDYLLNLKAKINDCNVDEAHVKITKLSLDLKAIDDKIDLKFNPDFAIKK 240
QY 298 LINDDTKKDMLGKLLSTGLVQNPNTIISKLTGKFDMLNISOHCVKQKQCPNSGCFR 357
Db 241 LINDDTKKDMLGKLLSTGLVQNPNTIISKLTGKFDMLNISOHCVKQKQCPNSGCFR 300
QY 358 HLDERECKLLNYKQEGDKCVENPNTCNENNGCGDADATCTEEDSGSRKKITCECTK 417
Db 301 HLDERECKLLNYKQEGDKCVENPNTCNENNGCGDADATCTEEDSGSRKKITCECTK 360
QY 418 PDSYPLFDGIFCS 430
Db 361 PDSYPLFDGIFCS 373

RESULT 9
Q25970
ID Q25970 PRELIMINARY; PRT; 569 AA.
AC Q25970;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NAR-2002 (TREMBlrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSPI.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutiwes S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSPI) of Plasmodium falciparum from
RT field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13347; BAA02608.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64502 MW; 4D761FF472753142 CRC64;

Query Match 84.9%; Score 1931; DB 5; Length 569;
Best Local Similarity 94.7%; Pred. No. 2e-86;
Matches 373; Conservative 7; Mismatches 8; Indels 6; Gaps 2;

QY 38 TDDDDKAMADIGSIEGRTMAISVTMDNLISGFENEYDVIIYKPLAGVYRSLLKKQIEKNI 97
Db 162 SEDNDYLDQVVTGE-----AISVTMDNLISGFENEYDVIIYKPLAGVYRSLLKKQIEKNI 216
QY 98 FTFNLNDLILNSRLKRRYFDLVLESOLMQPKHSSNEYYIETDSFKLNSQKNTLLKS 157
Db 217 IFNLNDLILNSRLKRRYFDLVLESOLMQPKHSSNEYYIETDSFKLNSQKNTLLKS 276
QY 158 YKIKESVENDIKFAOEGISYIEKVLAKYKDDLESIKKVIKEKEKFPSPPTTPSPAK 217
Db 277 YKIKESVENDIKFAOEGISYIEKVLAKYKDDLESIKKVIKEKEKFPSPPTTPSPAK 335
QY 218 TDQKESKFLPLPLTNIETLLNLVKNKIDDYLLNLKAKINDCNVDEAHVKITKLSLDL 277
Db 336 TDEQKESKFLPLPLTNIETLLNLVKNKIDDYLLNLKAKINDCNVDEAHVKITKLSLDL 395
QY 278 AIDDKIDLKFNPDFAIKKINDDTKKDMLGKLLSTGLVQNPNTIISKLTGKFDML 337
Db 396 AIDDKIDLKFNPDFAIKKINDDTKKDMLGKLLSTGLVQNPNTIISKLTGKFDML 455
QY 338 NISQHCQVKKQCPNSGCFRHLDERECKCLLNYKQEGDKCVENPNTCNENNGCGDADA 397
Db 456 NISQHCQVKKQCPNSGCFRHLDERECKCLLNYKQEGDKCVENPNTCNENNGCGDADA 515
QY 398 TCTEEDSGSRKKITCECTKPDSPYPLFDGIFCSS 431
Db 516 TCTEEDSGSRKKITCECTKPDSPYPLFDGIFCSS 549

RESULT 10
Q25980
ID Q25980 PRELIMINARY; PRT; 569 AA.
AC Q25980;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NAR-2002 (TREMBlrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSPI.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutiwes S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSPI) of Plasmodium falciparum from
RT field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13353; BAA02614.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64536 MW; 8008861DECECD8DC CRC64;

Query Match 84.9%; Score 1931; DB 5; Length 569;
Best Local Similarity 94.7%; Pred. No. 2e-86;
Matches 373; Conservative 7; Mismatches 8; Indels 6; Gaps 2;

QY 38 TDDDDKAMADIGSIEGRTMAISVTMDNLISGFENEYDVIIYKPLAGVYRSLLKKQIEKNI 97
Db 162 SEDNDYLDQVVTGE-----AISVTMDNLISGFENEYDVIIYKPLAGVYRSLLKKQIEKNI 216
QY 98 FTFNLNDLILNSRLKRRYFDLVLESOLMQPKHSSNEYYIETDSFKLNSQKNTLLKS 157
Db 217 IFNLNDLILNSRLKRRYFDLVLESOLMQPKHSSNEYYIETDSFKLNSQKNTLLKS 276
QY 158 YKIKESVENDIKFAOEGISYIEKVLAKYKDDLESIKKVIKEKEKFPSPPTTPSPAK 217
Db 277 YKIKESVENDIKFAOEGISYIEKVLAKYKDDLESIKKVIKEKEKFPSPPTTPSPAK 335
QY 218 TDQKESKFLPLPLTNIETLLNLVKNKIDDYLLNLKAKINDCNVDEAHVKITKLSLDL 277
Db 336 TDEQKESKFLPLPLTNIETLLNLVKNKIDDYLLNLKAKINDCNVDEAHVKITKLSLDL 395
QY 278 AIDDKIDLKFNPDFAIKKINDDTKKDMLGKLLSTGLVQNPNTIISKLTGKFDML 337
Db 396 AIDDKIDLKFNPDFAIKKINDDTKKDMLGKLLSTGLVQNPNTIISKLTGKFDML 455
QY 338 NISQHCQVKKQCPNSGCFRHLDERECKCLLNYKQEGDKCVENPNTCNENNGCGDADA 397
Db 456 NISQHCQVKKQCPNSGCFRHLDERECKCLLNYKQEGDKCVENPNTCNENNGCGDADA 515
QY 398 TCTEEDSGSRKKITCECTKPDSPYPLFDGIFCSS 431
Db 516 TCTEEDSGSRKKITCECTKPDSPYPLFDGIFCSS 549
```

```
RESULT 11
Q25982 ID Q25982 PRELIMINARY; PRT; 569 AA.
AC Q25982;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
DE MSP1.
GN Plasmodium falciparum.
OS Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OC NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL: D13354; BAA02615.1; -.
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64530 MW; 4D734580516D2142 CRC64;

Query Match 84.98; Score 1931; DB 5; Length 569;
Best Local Similarity 94.78; Pred. No. 2e-86;
Matches 373; Conservative 7; Mismatches 8; Indels 6; Gaps 2;

Qy 38 TDDDDKADIGSGREGTMAISVTMDNILSGFENEYDVYILKPLAGYVRSILKQIEKNI 97
Db 162 SEDNDEYLDQVTCG-----AISVTMDNILSGFENEYDVYILKPLAGYVRSILKQIEKNI 216
Qy 98 FTFNILNDILNSRLKKRYFDLVLESQNFHSSNEYIIDSFKLLNSEQNTLLKS 157
Db 217 ITFNILNDILNSRLKKRYFDLVLESQNFHSSNEYIIDSFKLLNSEQNTLLKS 276
Qy 158 YKYIKESVENDIKFAQEGISYEVYKLVAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAK 217
Db 277 YKYIKESVENDIKFAQEGISYEVYKLVAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAK 335
Qy 218 TDEQKESKFLPFLTNITETLNILNVLNKKIDYILNKKAKINDCNVEKDEAHVKITKLSDLK 277
Db 336 TDEQKESKFLPFLTNITETLNILNVLNKKIDYILNKKAKINDCNVEKDEAHVKITKLSDLK 395
Qy 278 AIDDKIDLFKNPYDFEAIKKLINDDTKMDLGLKLLSTGLVQNFNTIISKIEGKFDQML 337
Db 396 AIDDKIDLFKNPYDFEAIKKLINDDTKMDLGLKLLSTGLVQNFNTIISKIEGKFDQML 455
Qy 338 NISQHCYKVKQCPENSGCFRHLDERECKLLNKKQEGDKCVENPPTCNENNGCCADA 397
Db 456 NISQHCYKVKQCPENSGCFRHLDERECKLLNKKQEGDKCVENPPTCNENNGCCADA 515
Qy 398 TCTEEDSGSSRRKKTCTCKPDSYPLFDGIFCSS 431
Db 516 TCTEEDSGSSRRKKTCTCKPDSYPLFDGIFCSS 549

RESULT 12
Q43996 ID Q43996 PRELIMINARY; PRT; 373 AA.
AC Q43996;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DE Major merozoite surface protein (Fragment).
DE MSP-1.
GN Plasmodium falciparum.
OS Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
```

```
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=KENYA-2;
RA Qari S.H., Shi Y.P., Goldman I.F., Nahlen B.L., Tibayrenc M.,
RA Lal A.A.;
RT "Predicted and observed alleles of plasmodium falciparum merozoite
RT surface protein-1 (MSP-1), a potential malaria vaccine antigen.";
RL Mol. Biochem. Parasitol. 92:241-252(1998).
DR EMBL: AF040568; AAC39098.1; -.
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
FT NON_TER 1
FT NON_TER 373
SQ SEQUENCE 373 AA; 42902 MW; 86CD4B721E605A5F CRC64;

Query Match 84.88; Score 1929; DB 5; Length 373;
Best Local Similarity 98.98; Pred. No. 1.6e-86;
Matches 369; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 58 AISVTMDNILSGFENEYDVYILKPLAGYVRSILKQIEKNIETFNILNDILNSRLKKRY 117
Db 1 AISVTMDNILSGFENEYDVYILKPLAGYVRSILKQIEKNIETFNILNDILNSRLKKRY 60
Qy 118 FLDVLESQNFHSSNEYIIDSFKLLNSEQNTLLKSFKLLNSEQNTLLKSFKLLNSEQNTLLKS 177
Db 61 FLDVLESQNFHSSNEYIIDSFKLLNSEQNTLLKSFKLLNSEQNTLLKSFKLLNSEQNTLLKS 120
Qy 178 YYEKVLAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAKTDEQKESKFLPFLTNITL 237
Db 121 YYEKVLAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAKTDEQKESKFLPFLTNITL 180
Qy 238 YNNLVNKIDDDYILNKKAKINDCNVEKDEAHVKITKLSDLKAIIDKIDLFKNPYDFEAIKK 297
Db 181 YNNLVNKIDDDYILNKKAKINDCNVEKDEAHVKITKLSDLKAIIDKIDLFKNPYDFEAIKK 240
Qy 298 LINDDTKMDLGLKLLSTGLVQNFNTIISKIEGKFDQMLNISQHCYKVKQCPENSGCFR 357
Db 241 LINDDTKMDLGLKLLSTGLVQNFNTIISKIEGKFDQMLNISQHCYKVKQCPENSGCFR 300
Qy 358 HLDRECECKLLNKKQEGDKCVENPPTCNENNGCCADATCTEEDSGSSRRKKTCTCK 417
Db 301 HLDRECECKLLNKKQEGDKCVENPPTCNENNGCCADATCTEEDSGSSRRKKTCTCK 360
Qy 418 PDSYPLFDGIFCS 430
Db 361 SDSYPLFDGIFCS 373

RESULT 13
Q25728 ID Q25728 PRELIMINARY; PRT; 373 AA.
AC Q25728;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1, 42 kDa C-terminal region (Fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=INFECTED HUMAN BLOOD SAMPLE;
RA Shi Y.-P., Alpers M.P., Pova M.M., Nahlen B.L., Oloo A.G., Lal A.A.;
RT "Sequence of the C-terminal region of merozoite surface protein-1
RT (MSP-1) in field-derived plasmodium falciparum."
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: U20733; AAA62224.1; -.
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
FT NON_TER 1
FT NON_TER 1
```

```
FT NON_TER 373 373
SQ SEQUENCE 373 AA: 42815 MW: 9CF4DDAF38CC4054 CRC64;

Query Match 84.7%; Score 1928; DB 5; Length 373;
Best Local Similarity 98.7%; Pred. No. 1.8e-86;
Matches 368; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 58 AISVTMDNLISGFENEYDVYILKPLAGVYRSLLKQIEKNIFTNPNLNDILNSRLKRRKY 117
Db 1 AISVTMDNLISGFENEYDVYILKPLAGVYRSLLKQIEKNIFTNPNLNDILNSRLKRRKY 60

QY 118 FLDVLESMLQPKHSSNEYIIEDSFKLLNSEQKNTLLSKYIKESVENDIKFAQEGIS 177
Db 61 FLDVLESMLQPKHSSNEYIIEDSFKLLNSEQKNTLLSKYIKESVENDIKFAQEGIS 120

QY 178 YVEKVLAKYKDDLESIKVKIKEEKEKFPSSPTTTPSPAKTDEQKESKFLPFLTNIELT 237
Db 121 YVEKVLAKYKDDLESIKVKIKEEKEKFPSSPTTTPSPAKTDEQKESKFLPFLTNIELT 180

QY 238 YNNLVNKIDDYILINLAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 297
Db 181 YNNLVNKIDDYILINLAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240

QY 298 LINDDTKKDMLGKLLSTGLVQNFPTIISKLEGFQDMLNISOHCYKVKQCPENSGCGR 357
Db 241 LINDDTKKDMLGKLLSTGLVQNFPTIISKLEGFQDMLNISOHCYKVKQCPENSGCGR 300

QY 358 HLDERECKCLLNYKQEGDKCVENPNTCNENNGGCDADATCTEEDSGSSRKKITCECTK 417
Db 301 HLDERECKCLLNYKQEGDKCVENPNTCNENNGGCDADATCTEEDSGSSRKKITCECTK 360

QY 418 PDSYPLFDGIFCS 430
Db 361 PDSYPLFDGIFCS 373

RESULT 14
Q25724 PRELIMINARY; PRT: 373 AA.
AC Q25724;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE Merozoite surface protein 1, 42 kDa C-terminal region (Fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-INFECTED HUMAN BLOOD SAMPLE;
RA Shi Y.-P., Alpers M.P., Povoas M.M., Nahlen B.L., Oloo A.G., Lal A.A.;
RT "Sequence of the C-terminal region of merozoite surface protein-1
RT (MSP-1) in field-derived Plasmodium falciparum.";
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U20729; AAA62220.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
FT NON_TER 373 373
FT SEQUENCE 373 AA: 42839 MW: A21A91B2076F415A CRC64;

Query Match 84.6%; Score 1924; DB 5; Length 373;
Best Local Similarity 98.4%; Pred. No. 2.8e-86;
Matches 367; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 58 AISVTMDNLISGFENEYDVYILKPLAGVYRSLLKQIEKNIFTNPNLNDILNSRLKRRKY 117
Db 1 AISVTMDNLISGFENEYDVYILKPLAGVYRSLLKQIEKNIFTNPNLNDILNSRLKRRKY 60

QY 118 FLDVLESMLQPKHSSNEYIIEDSFKLLNSEQKNTLLSKYIKESVENDIKFAQEGIS 177
Db 61 FLDVLESMLQPKHSSNEYIIEDSFKLLNSEQKNTLLSKYIKESVENDIKFAQEGIS 120
```

```
QY 178 YVEKVLAKYKDDLESIKVKIKEEKEKFPSSPTTTPSPAKTDEQKESKFLPFLTNIELT 237
Db 121 YVEKVLAKYKDDLESIKVKIKEEKEKFPSSPTTTPSPAKTDEQKESKFLPFLTNIELT 180

QY 238 YNNLVNKIDDYILINLAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 297
Db 181 YNNLVNKIDDYILINLAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240

QY 298 LINDDTKKDMLGKLLSTGLVQNFPTIISKLEGFQDMLNISOHCYKVKQCPENSGCGR 357
Db 241 LINDDTKKDMLGKLLSTGLVQNFPTIISKLEGFQDMLNISOHCYKVKQCPENSGCGR 300

QY 358 HLDERECKCLLNYKQEGDKCVENPNTCNENNGGCDADATCTEEDSGSSRKKITCECTK 417
Db 301 HLDERECKCLLNYKQEGDKCVENPNTCNENNGGCDADATCTEEDSGSSRKKITCECTK 360

QY 418 PDSYPLFDGIFCS 430
Db 361 PDSYPLFDGIFCS 373

RESULT 15
Q25717 PRELIMINARY; PRT: 372 AA.
AC Q25717;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE Merozoite surface protein 1, 42 kDa C-terminal region (Fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-INFECTED HUMAN BLOOD SAMPLE;
RA Shi Y.-P., Alpers M.P., Povoas M.M., Nahlen B.L., Oloo A.G., Lal A.A.;
RT "Sequence of the C-terminal region of merozoite surface protein-1
RT (MSP-1) in field-derived Plasmodium falciparum.";
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U20653; AAA62213.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
FT NON_TER 372 372
FT SEQUENCE 372 AA: 42687 MW: 3F2BF1152598FB10 CRC64;

Query Match 84.5%; Score 1922.5; DB 5; Length 372;
Best Local Similarity 98.9%; Pred. No. 3.3e-86;
Matches 369; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 58 AISVTMDNLISGFENEYDVYILKPLAGVYRSLLKQIEKNIFTNPNLNDILNSRLKRRKY 117
Db 1 AISVTMDNLISGFENEYDVYILKPLAGVYRSLLKQIEKNIFTNPNLNDILNSRLKRRKY 60

QY 118 FLDVLESMLQPKHSSNEYIIEDSFKLLNSEQKNTLLSKYIKESVENDIKFAQEGIS 177
Db 61 FLDVLESMLQPKHSSNEYIIEDSFKLLNSEQKNTLLSKYIKESVENDIKFAQEGIS 120

QY 178 YVEKVLAKYKDDLESIKVKIKEEKEKFPSSPTTTPSPAKTDEQKESKFLPFLTNIELT 237
Db 121 YVEKVLAKYKDDLESIKVKIKEEKEKFPSSPTTTPSPAKTDEQKESKFLPFLTNIELT 179

QY 238 YNNLVNKIDDYILINLAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 297
Db 180 YNNLVNKIDDYILINLAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 239

QY 298 LINDDTKKDMLGKLLSTGLVQNFPTIISKLEGFQDMLNISOHCYKVKQCPENSGCGR 357
Db 240 LINDDTKKDMLGKLLSTGLVQNFPTIISKLEGFQDMLNISOHCYKVKQCPENSGCGR 299

QY 358 HLDERECKCLLNYKQEGDKCVENPNTCNENNGGCDADATCTEEDSGSSRKKITCECTK 417
```

```
Db 300 HLDRECKLLNYKQEGDKCVENPNTCNENGGCCADATCTEDSGSRKKITCCTK 359
Qy 418 PDSYPLFDGIFCS 430
Db 360 PDSYPLFDGIFCS 372

RESULT 16
Q25983 ID Q25983 PRELIMINARY; PRT; 569 AA.
AC Q25983;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwatitwies S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL: D13355; BAA02616.1;
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
FT SEQUENCE 569 AA; 64491 MW; 57AGB62FF72CE885 CRC64;

Query Match 84.4%; Score 1921; DB 5; Length 569;
Best Local Similarity 94.2%; Pred. No. 6.1e-86;
Matches 371; Conservative 8; Mismatches 9; Indels 6; Gaps 2;

Qy 38 TODDKAMADIGTEGRGTMAISVTMDNILSGFENEDVIYLPAGVYSLKQIEKNI 97
Db 162 SEDNEVDQVVTGE-----AISTMDNILSGFENEDVIYLPAGVYSLKQIEKNI 216
Qy 98 FTFNLNDILNSRLKRRKRYFLDVLSDLMQFKHISSENYIIEDSFKLLNSEQKNTLLKS 157
Db 217 FTFNLNDILNSRLKRRKRYFLDVLSDLMQFKHISSENYIIEDSFKLLNSEQKNTLLKS 276
Qy 158 YKIKESVENDIKFAQEGISYKVLAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAK 217
Db 277 YKIKESVENDIKFAQEGISYKVLAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAK 335
Qy 218 TDEQKESKFLPFLTNTETLYNNLVNKIDDIYLNKAKINDCNVKEKDEAHVKITKLSDLK 277
Db 336 TDEQKESKFLPFLTNTETLYNNLVNKIDDIYLNKAKINDCNVKEKDEAHVKITKLSDLK 395
Qy 278 AIDDKIDLFKNPDVFEAKKILNDTKKMLGKLLSTGLVQNFNTIISKIEGKFODML 337
Db 396 AIDDKIDLFKNPDVFEAKKILNDTKKMLGKLLSTGLVQNFNTIISKIEGKFODML 455
Qy 338 NISOHCVKKQCPNSGCCFPHLDRECKCLLYNKQEGDKCVENPNTCNENGGCCADA 397
Db 456 NISOHCVKKQCPNSGCCFPHLDRECKCLLYNKQEGDKCVENPNTCNENGGCCADA 515
Qy 398 TCTEEDSGSRKKITCCTKPDYPLFDGIFCS 431
Db 516 KCTEEDSGSGNGKITCCTKPDYPLFDGIFCS 549

RESULT 17
Q43995 ID Q43995 PRELIMINARY; PRT; 373 AA.
AC Q43995;
DT 01-JUN-1998 (Tremblrel. 06, Created)
```

```
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Merozoite surface protein-1 (Fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=KENYA-1.
RX MEDLINE=98319411; PubMed=9657329;
RA Qari S.H., Shi Y.P., Goldman I.F., Nahlen B.L., Tibayrenc M.,
RA Lal A.A.;
RT "Predicted and observed alleles of Plasmodium falciparum merozoite
RT surface protein-1 (MSP-1), a potential malaria vaccine antigen."
RL Mol. Biochem. Parasitol. 92:241-252(1998).
DR EMBL: AF040567; AAC39097.1;
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
FT NON_TER 1
FT SEQUENCE 373 AA; 42873 MW; 886CF169A7AF5194 CRC64;

Query Match 84.4%; Score 1919; DB 5; Length 373;
Best Local Similarity 98.7%; Pred. No. 4.9e-86;
Matches 368; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 58 AISTMDNILSGFENEDVIYLPAGVYSLKQIEKNIETFNLDILNSRLKRRKY 117
Db 1 AISTMDNILSGFENEDVIYLPAGVYSLKQIEKNIETFNLDILNSRLKRRKY 60
Qy 118 FLDVLESDLMQFKHISSENYIIEDSFKLLNSEQKNTLLKSYKIKESVENDIKFAQEGIS 177
Db 61 FLDVLESDLMQFKHISSENYIIEDSFKLLNSEQKNTLLKSYKIKESVENDIKFAQEGIS 120
Qy 178 YKVLAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAKTDEQKESKFLPFLTNIETL 237
Db 121 YKVLAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAKTDEQKESKFLPFLTNIETL 180
Qy 238 YNNLVNKIDDIYLNKAKINDCNVKEKDEAHVKITKLSDLKAIIDKIDLFKNPDVFEAKK 297
Db 181 YNNLVNKIDDIYLNKAKINDCNVKEKDEAHVKITKLSDLKAIIDKIDLFKNPDVFEAKK 240
Qy 298 LINDTKKMLGKLLSTGLVQNFNTIISKIEGKFODMLNISOHCVKKQCPNSGCCFR 357
Db 241 LINDTKKMLGKLLSTGLVQNFNTIISKIEGKFODMLNISOHCVKKQCPNSGCCFR 300
Qy 358 HLDRECKCLLYNKQEGDKCVENPNTCNENGGCCADATCTEDSGSRKKITCCTK 417
Db 301 HLDRECKCLLYNKQEGDKCVENPNTCNENGGCCADATCTEDSGSRKKITCCTK 360
Qy 418 PDSYPLFDGIFCS 430
Db 361 PDSYPLFDGIFCS 373

RESULT 18
Q25721 ID Q25721 PRELIMINARY; PRT; 373 AA.
AC Q25721;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Merozoite surface protein 1, 42 kDa C-terminal region (Fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=INFECTED HUMAN BLOOD SAMPLE;
RA Shi Y.-P., Alpers M.P., Pova M.M., Nahlen B.L., Oloo A.G., Lal A.A.;
RA "Sequence of the C-terminal region of merozoite surface protein-1
```

RT (MSP-1) in field-derived Plasmodium falciparum.";  
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U20726; AAA62217.1; -;  
DR InterPro: IPR000561; EGF-like.  
DR Pfam: PF00008; EGF; 1;  
FT NON\_TER 1 373  
SQ SEQUENCE 373 AA; 42848 MW; EE0700233D7F4D4E CRC64;

Query Match 84.4%; Score 1919; DB 5; Length 373;  
Best Local Similarity 98.1%; Pred. No. 4.9e-86;  
Matches 366; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 58 A1SVTMDN1LSGFENEYDVYILKPLAGVYRSLKQIEKNIFTN1LNLDILNSRLKRRKY 117  
Db 1 A1SVTMDN1LSGFENEYDVYILKPLAGVYRSLKQIEKNIFTN1LNLDILNSRLKRRKY 60  
QY 118 FLDVLESOLMQFKHSSNEYIIEDSFKLLNSEQKNTLLSKYIKESVENDIKFAQEGIS 177  
Db 61 FLDVLESOLMQFKHSSNEYIIEDSFKLLNSEQKNTLLSKYIKESVENDIKFAQEGIS 120  
QY 178 YVEKVLAKYKDDLESIKKVIKEEKFPSSPTTPSPAKTDEQKESKFLPLTNIELT 237  
Db 121 YVEKVLAKYKDDLESIKKVIKEEKFPSSPTTPSPAKTDEQKESKFLPLTNIELT 180  
QY 238 YNNLVNKIDYILN1LAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFPKNPYDFEAIKK 297  
Db 181 YNNLVNKIDYILN1LAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFPKNPYDFEAIKK 240  
QY 298 LINDTKDKMLGKLLSTGLVQNPFTIISKLEGGFQDMLN1SOHCVKKQCPENSGCPR 357  
Db 241 LINDTKDKMLGKLLSTGLVQNPFTIISKLEGGFQDMLN1SOHCVKKQCPENSGCPR 300  
QY 358 HLDERECKCLLNYKQEGDKCVENPNTCNENNGGCDADATCTEEDSGSRKKITCECTK 417  
Db 301 HLDERECKCLLNYKQEGDKCVENPNTCNENNGGCDADATCTEEDSGSGNKKVCECTK 360  
QY 418 PDSYPLFDGIFCS 430  
Db 361 PDSYPLFDGIFCS 373

RESULT 19  
Q25718 PRELIMINARY; PRT; 372 AA.  
AC Q25718;  
DT 01-NOV-1996 (TReMBLrel. 01, Created)  
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE Merozoite surface protein 1, 42 kDa C-terminal region (Fragment).  
GN MSP-1.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-INFECTED HUMAN BLOOD SAMPLE;  
RA Shi Y.-P., Alpers M.P., Povoa M.M., Nahlen B.L., Oloo A.G., Lal A.A.;  
RT "Sequence of the C-terminal region of merozoite surface protein-1  
RT (MSP-1) in field-derived Plasmodium falciparum.";  
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U20654; AAA62214.1; -;  
DR InterPro: IPR000561; EGF-like.  
DR Pfam: PF00008; EGF; 1;  
FT NON\_TER 1 372  
SQ SEQUENCE 372 AA; 42660 MW; E9F088252598FB0B CRC64;

Query Match 84.3%; Score 1917.5; DB 5; Length 372;  
Best Local Similarity 98.7%; Pred. No. 5.8e-86;  
Matches 368; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 58 A1SVTMDN1LSGFENEYDVYILKPLAGVYRSLKQIEKNIFTN1LNLDILNSRLKRRKY 117

Db 1 A1SVTMDN1LSGFENEYDVYILKPLAGVYRSLKQIEKNIFTN1LNLDILNSRLKRRKY 60  
QY 118 FLDVLESOLMQFKHSSNEYIIEDSFKLLNSEQKNTLLSKYIKESVENDIKFAQEGIS 177  
Db 61 FLDVLESOLMQFKHSSNEYIIEDSFKLLNSEQKNTLLSKYIKESVENDIKFAQEGIS 120  
QY 178 YVEKVLAKYKDDLESIKKVIKEEKFPSSPTTPSPAKTDEQKESKFLPLTNIELT 237  
Db 121 YVEKVLAKYKDDLESIKKVIKEEKFPSSPTTPSPAKTDEQKESKFLPLTNIELT 179  
QY 238 YNNLVNKIDYILN1LAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFPKNPYDFEAIKK 297  
Db 180 YNNLVNKIDYILN1LAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFPKNPYDFEAIKK 239  
QY 298 LINDTKDKMLGKLLSTGLVQNPFTIISKLEGGFQDMLN1SOHCVKKQCPENSGCPR 357  
Db 240 LINDTKDKMLGKLLSTGLVQNPFTIISKLEGGFQDMLN1SOHCVKKQCPENSGCPR 299  
QY 358 HLDERECKCLLNYKQEGDKCVENPNTCNENNGGCDADATCTEEDSGSRKKITCECTK 417  
Db 300 HLDERECKCLLNYKQEGDKCVENPNTCNENNGGCDADATCTEEDSGSGSRKKITCECTK 359  
QY 418 PDSYPLFDGIFCS 430  
Db 360 PDSYPLFDGIFCS 372

RESULT 20  
Q25719 PRELIMINARY; PRT; 372 AA.  
AC Q25719;  
DT 01-NOV-1996 (TReMBLrel. 01, Created)  
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE Merozoite surface protein 1, 42 kDa C-terminal region (Fragment).  
GN MSP-1.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-INFECTED HUMAN BLOOD SAMPLE;  
RA Shi Y.-P., Alpers M.P., Povoa M.M., Nahlen B.L., Oloo A.G., Lal A.A.;  
RT "Sequence of the C-terminal region of merozoite surface protein-1  
RT (MSP-1) in field-derived Plasmodium falciparum.";  
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U20655; AAA62215.1; -;  
DR InterPro: IPR000561; EGF-like.  
DR Pfam: PF00008; EGF; 1;  
FT NON\_TER 1 372  
SQ SEQUENCE 372 AA; 42686 MW; 583A33972DB05FB4 CRC64;

Query Match 84.3%; Score 1917.5; DB 5; Length 372;  
Best Local Similarity 98.7%; Pred. No. 5.8e-86;  
Matches 368; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 58 A1SVTMDN1LSGFENEYDVYILKPLAGVYRSLKQIEKNIFTN1LNLDILNSRLKRRKY 117  
Db 1 A1SVTMDN1LSGFENEYDVYILKPLAGVYRSLKQIEKNIFTN1LNLDILNSRLKRRKY 60  
QY 118 FLDVLESOLMQFKHSSNEYIIEDSFKLLNSEQKNTLLSKYIKESVENDIKFAQEGIS 177  
Db 61 FLDVLESOLMQFKHSSNEYIIEDSFKLLNSEQKNTLLSKYIKESVENDIKFAQEGIS 120  
QY 178 YVEKVLAKYKDDLESIKKVIKEEKFPSSPTTPSPAKTDEQKESKFLPLTNIELT 237  
Db 121 YVEKVLAKYKDDLESIKKVIKEEKFPSSPTTPSPAKTDEQKESKFLPLTNIELT 179  
QY 238 YNNLVNKIDYILN1LAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFPKNPYDFEAIKK 297  
Db 180 YNNLVNKIDYILN1LAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFPKNPYDFEAIKK 239





```
Db 162 SEDNDEYLDQVVTGE-----AISTVMDNILSGFENEYDVILKPLAGYVRSLLKKQIEKNI 216
QY 98 FTFLNLNDILNSRLKRRKYFLDVLSDLMQFKHISSEYIIEDSFKLLNSEQKNTLLKS 157
Db 217 ITFNLNLNDILNSRLKRRKYFLDVLSDLMQFKHISSEYIIEDSFKLLNSEQKNTLLKS 276
QY 158 YKYIKESVENDIKFAQEGISYVERVLAKYKDDLESIKKVIKEEKEKFPSSPTTTPSPAK 217
Db 277 YKYIKESVENDIKFAQEGISYVERVLAKYKDDLESIKKVIKEEKE-PPSSPTTTPSPAK 335
QY 218 TDEQKESKFLPFLTNIETLYNNLVNKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLK 277
Db 336 TDEQKESKFLPFLTNIETLYNNLVNKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLK 395
QY 278 AIDDKIDLFPKPYDFAIKKLINDTKKMLGKLLSTGLVONFPNTIISKLEIGKFDQML 337
Db 455 AIDDKIDLFPKPYDFAIKKLINDTKKMLGKLLSTGLVONFPNTIISKLEIGKFDQML 455
QY 338 NISQHCVKKQCPENSGCFRHLDEREECKLLNYKQEGDKCVENPNPTCNENNGCCDADA 397
Db 456 NISQHCVKKQCPENSGCFRHLDEREECKLLNYKQEGDKCVENPNPTCNENNGCCDADA 515
QY 398 TCTEEDSGSSRRKKTCTCTKPDSPFLPDGIFCSS 431
Db 516 KCTEEDSGSNGKKITCTCTKPDSPFLPDGIFCSS 549

RESULT 26
Q25977 ID Q25977 PRELIMINARY; PRT; 569 AA.
AC Q25977;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL: D13350; BAA02611.1; -
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64485 MW; A26C45914AF01396 CRC64;

Query Match 84.28; Score 1915; DB 5; Length 569;
Best Local Similarity 93.98; Pred. No. 1.2e-85;
Matches 370; Conservative 8; Mismatches 10; Indels 6; Gaps 2;

QY 38 TDDDDKAMADTSGTGRTMAISVTMDNILSGFENEYDVILKPLAGYVRSLLKKQIEKNI 97
Db 162 SEDNDEYLDQVVTGE-----AISTVMDNILSGFENEYDVILKPLAGYVRSLLKKQIEKNI 216
QY 98 FTFLNLNDILNSRLKRRKYFLDVLSDLMQFKHISSEYIIEDSFKLLNSEQKNTLLKS 157
Db 217 ITFNLNLNDILNSRLKRRKYFLDVLSDLMQFKHISSEYIIEDSFKLLNSEQKNTLLKS 276
QY 158 YKYIKESVENDIKFAQEGISYVERVLAKYKDDLESIKKVIKEEKEKFPSSPTTTPSPAK 217
Db 277 YKYIKESVENDIKFAQEGISYVERVLAKYKDDLESIKKVIKEEKE-PPSSPTTTPSPAK 335
QY 218 TDEQKESKFLPFLTNIETLYNNLVNKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLK 277
Db 336 TDEQKESKFLPFLTNIETLYNNLVNKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLK 395
QY 278 AIDDKIDLFPKPYDFAIKKLINDTKKMLGKLLSTGLVONFPNTIISKLEIGKFDQML 337
Db 455 AIDDKIDLFPKPYDFAIKKLINDTKKMLGKLLSTGLVONFPNTIISKLEIGKFDQML 455
QY 338 NISQHCVKKQCPENSGCFRHLDEREECKLLNYKQEGDKCVENPNPTCNENNGCCDADA 397
Db 456 NISQHCVKKQCPENSGCFRHLDEREECKLLNYKQEGDKCVENPNPTCNENNGCCDADA 515
QY 398 TCTEEDSGSSRRKKTCTCTKPDSPFLPDGIFCSS 431
Db 516 KCTEEDSGSNGKKITCTCTKPDSPFLPDGIFCSS 549

RESULT 26
Q25977 ID Q25977 PRELIMINARY; PRT; 569 AA.
AC Q25977;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL: D13350; BAA02611.1; -
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64485 MW; A26C45914AF01396 CRC64;

Query Match 84.28; Score 1915; DB 5; Length 569;
Best Local Similarity 93.98; Pred. No. 1.2e-85;
Matches 370; Conservative 8; Mismatches 10; Indels 6; Gaps 2;

QY 38 TDDDDKAMADTSGTGRTMAISVTMDNILSGFENEYDVILKPLAGYVRSLLKKQIEKNI 97
Db 162 SEDNDEYLDQVVTGE-----AISTVMDNILSGFENEYDVILKPLAGYVRSLLKKQIEKNI 216
QY 98 FTFLNLNDILNSRLKRRKYFLDVLSDLMQFKHISSEYIIEDSFKLLNSEQKNTLLKS 157
Db 217 ITFNLNLNDILNSRLKRRKYFLDVLSDLMQFKHISSEYIIEDSFKLLNSEQKNTLLKS 276
QY 158 YKYIKESVENDIKFAQEGISYVERVLAKYKDDLESIKKVIKEEKEKFPSSPTTTPSPAK 217
Db 277 YKYIKESVENDIKFAQEGISYVERVLAKYKDDLESIKKVIKEEKE-PPSSPTTTPSPAK 335
QY 218 TDEQKESKFLPFLTNIETLYNNLVNKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLK 277
```

```
Db 336 TDEQKESKFLPFLTNIETLYNNLVNKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLK 395
QY 278 AIDDKIDLFPKPYDFAIKKLINDTKKMLGKLLSTGLVONFPNTIISKLEIGKFDQML 337
Db 396 AIDDKIDLFPKPYDFAIKKLINDTKKMLGKLLSTGLVONFPNTIISKLEIGKFDQML 455
QY 338 NISQHCVKKQCPENSGCFRHLDEREECKLLNYKQEGDKCVENPNPTCNENNGCCDADA 397
Db 456 NISQHCVKKQCPENSGCFRHLDEREECKLLNYKQEGDKCVENPNPTCNENNGCCDADA 515
QY 398 TCTEEDSGSSRRKKTCTCTKPDSPFLPDGIFCSS 431
Db 516 KCTEEDSGSNGKKITCTCTKPDSPFLPDGIFCSS 549

RESULT 27
Q25979 ID Q25979 PRELIMINARY; PRT; 569 AA.
AC Q25979;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL: D13352; BAA02613.1; -
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64457 MW; A2691FE569E80396 CRC64;

Query Match 84.28; Score 1915; DB 5; Length 569;
Best Local Similarity 93.98; Pred. No. 1.2e-85;
Matches 370; Conservative 8; Mismatches 10; Indels 6; Gaps 2;

QY 38 TDDDDKAMADTSGTGRTMAISVTMDNILSGFENEYDVILKPLAGYVRSLLKKQIEKNI 97
Db 162 SEDNDEYLDQVVTGE-----AISTVMDNILSGFENEYDVILKPLAGYVRSLLKKQIEKNI 216
QY 98 FTFLNLNDILNSRLKRRKYFLDVLSDLMQFKHISSEYIIEDSFKLLNSEQKNTLLKS 157
Db 217 ITFNLNLNDILNSRLKRRKYFLDVLSDLMQFKHISSEYIIEDSFKLLNSEQKNTLLKS 276
QY 158 YKYIKESVENDIKFAQEGISYVERVLAKYKDDLESIKKVIKEEKEKFPSSPTTTPSPAK 217
Db 277 YKYIKESVENDIKFAQEGISYVERVLAKYKDDLESIKKVIKEEKE-PPSSPTTTPSPAK 335
QY 218 TDEQKESKFLPFLTNIETLYNNLVNKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLK 277
Db 336 TDEQKESKFLPFLTNIETLYNNLVNKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLK 395
QY 278 AIDDKIDLFPKPYDFAIKKLINDTKKMLGKLLSTGLVONFPNTIISKLEIGKFDQML 337
Db 396 AIDDKIDLFPKPYDFAIKKLINDTKKMLGKLLSTGLVONFPNTIISKLEIGKFDQML 455
QY 338 NISQHCVKKQCPENSGCFRHLDEREECKLLNYKQEGDKCVENPNPTCNENNGCCDADA 397
Db 456 NISQHCVKKQCPENSGCFRHLDEREECKLLNYKQEGDKCVENPNPTCNENNGCCDADA 515
QY 398 TCTEEDSGSSRRKKTCTCTKPDSPFLPDGIFCSS 431
Db 516 KCTEEDSGSNGKKITCTCTKPDSPFLPDGIFCSS 549
```



```
RESULT 28
Q9NHX1 PRELIMINARY; PRT; 1694 AA.
AC Q9NHX1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Major merozoite surface antigen.
GN Gp195.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RC STRAIN=FCCL/HN;
RA Shan Z.X., Yu X.B., Li X.R., Ma C.L., Fang J.M.;
RT "Molecular cloning and sequence analysis of major merozoite surface
antigen(gp195)gene of Plasmodium falciparum isolate FCCL/HN.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF218248; AAF27526.1; -.
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
KW Merozoite.
SQ SEQUENCE 1694 AA; 192767 MW; B51634A49E0F6728 CRC64;

Query Match 84.2%; Score 1915; DB 5; Length 1694;
Best Local Similarity 93.9%; Pred. No. 3.8e-85;
Matches 370; Conservative 8; Mismatches 10; Indels 6; Gaps 2;

QY 38 TDDDDKAMADIGSIEGRGTMAISVTMDNILSGFENEYDVYILKPLAGVYRSLLKKQIEKNI 97
Db 1287 SEDNDEYLDQVVTGE-----AISVTMDNILSGFENEYDVYILKPLAGVYRSLLKKQIEKNI 1341
QY 98 FTENLNLNDILNSRLKRRKYFLDVLSDLMQFKHISSEYIIEDSFKLLNSEQKNTLLKS 157
Db 1342 ITENLNLNDILNSRLKRRKYFLDVLSDLMQFKHISSEYIIEDSFKLLNSEQKNTLLKS 1401
QY 158 YKIKESVENDIKFAQEGISYSEKVLAKYKDDLESIKKVIKEKEKFPSPPTTPSPAK 217
Db 1402 YKIKESVENDIKFAQEGISYSEKVLAKYKDDLESIKKVIKEKE-PPSSPPTTPSPAK 1460
QY 218 TDQKKESKFLPPLTNIETLYNNLVNKIDYDLINLAKINDCNVDEKDEAHVKITKLSDLK 277
Db 1461 TDQKKESKFLPPLTNIETLYNNLVNKIDYDLINLAKINDCNVDEKDEAHVKITKLSDLK 1520
QY 278 AIDDKIDLFPKNPYDFAIKKLINDDTKMDLGLSTGLVQNPNTIISKLEGGKFDML 337
Db 1521 AIDDKIDLFPKNPYDFAIKKLINDDTKMDLGLSTGLVQNPNTIISKLEGGKFDML 1580
QY 338 NISQHCVKKQCPENSGCFRHLDERECKLLNYKQEGDKCVENPNPTCNENNGGCDADA 397
Db 1581 NISQHCVKKQCPENSGCFRHLDERECKLLNYKQEGDKCVENPNPTCNENNGGCDADA 1640
QY 398 TCTEEDSGSRKKITCECTKPDSPYPLFDGIFCSS 431
Db 1641 KCTEEDSGSGNGKITCECTKPDSPYPLFDGIFCSS 1674

RESULT 29
Q9T2T5 PRELIMINARY; PRT; 1694 AA.
AC Q9T2T5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1.
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RC STRAIN=HN2;
RA Jiang G., Liu R., Daubenberger C.A., Pluschke G.;
RT "Sequence analysis of the MSP 1 gene of Plasmodium falciparum from
Hainan, China.";
RL Chung Kuo Chi Sheng Chung Hsueh Yu Chi Sheng Chung 17:294-297(1999).
DR EMBL: AF062349; AAC72885.1; -.
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
KW Merozoite surface protein 1.
SQ SEQUENCE 1694 AA; 193763 MW; 385526D0DA56FD1D CRC64;

Query Match 84.2%; Score 1915; DB 5; Length 1704;
Best Local Similarity 93.9%; Pred. No. 3.8e-85;
Matches 370; Conservative 8; Mismatches 10; Indels 6; Gaps 2;

QY 38 TDDDDKAMADIGSIEGRGTMAISVTMDNILSGFENEYDVYILKPLAGVYRSLLKKQIEKNI 97
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=HN1;
RA Jiang G., Liu R., Daubenberger C.A., Pluschke G.;
RT "Sequence analysis of the MSP 1 gene of Plasmodium falciparum from
Hainan, China.";
RL Chung Kuo Chi Sheng Chung Hsueh Yu Chi Sheng Chung 17:294-297(1999).
DR EMBL: AF062348; AAC72884.1; -.
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
SQ SEQUENCE 1694 AA; 192795 MW; 84CFC0E709F5673B CRC64;

Query Match 84.2%; Score 1915; DB 5; Length 1694;
Best Local Similarity 93.9%; Pred. No. 3.8e-85;
Matches 370; Conservative 8; Mismatches 10; Indels 6; Gaps 2;

QY 38 TDDDDKAMADIGSIEGRGTMAISVTMDNILSGFENEYDVYILKPLAGVYRSLLKKQIEKNI 97
Db 1287 SEDNDEYLDQVVTGE-----AISVTMDNILSGFENEYDVYILKPLAGVYRSLLKKQIEKNI 1341
QY 98 FTENLNLNDILNSRLKRRKYFLDVLSDLMQFKHISSEYIIEDSFKLLNSEQKNTLLKS 157
Db 1342 ITENLNLNDILNSRLKRRKYFLDVLSDLMQFKHISSEYIIEDSFKLLNSEQKNTLLKS 1401
QY 158 YKIKESVENDIKFAQEGISYSEKVLAKYKDDLESIKKVIKEKEKFPSPPTTPSPAK 217
Db 1402 YKIKESVENDIKFAQEGISYSEKVLAKYKDDLESIKKVIKEKE-PPSSPPTTPSPAK 1460
QY 218 TDQKKESKFLPPLTNIETLYNNLVNKIDYDLINLAKINDCNVDEKDEAHVKITKLSDLK 277
Db 1461 TDQKKESKFLPPLTNIETLYNNLVNKIDYDLINLAKINDCNVDEKDEAHVKITKLSDLK 1520
QY 278 AIDDKIDLFPKNPYDFAIKKLINDDTKMDLGLSTGLVQNPNTIISKLEGGKFDML 337
Db 1521 AIDDKIDLFPKNPYDFAIKKLINDDTKMDLGLSTGLVQNPNTIISKLEGGKFDML 1580
QY 338 NISQHCVKKQCPENSGCFRHLDERECKLLNYKQEGDKCVENPNPTCNENNGGCDADA 397
Db 1581 NISQHCVKKQCPENSGCFRHLDERECKLLNYKQEGDKCVENPNPTCNENNGGCDADA 1640
QY 398 TCTEEDSGSRKKITCECTKPDSPYPLFDGIFCSS 431
Db 1641 KCTEEDSGSGNGKITCECTKPDSPYPLFDGIFCSS 1674

RESULT 30
Q9T2T4 PRELIMINARY; PRT; 1704 AA.
AC Q9T2T4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1.
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RC STRAIN=HN2;
RA Jiang G., Liu R., Daubenberger C.A., Pluschke G.;
RT "Sequence analysis of the MSP 1 gene of Plasmodium falciparum from
Hainan, China.";
RL Chung Kuo Chi Sheng Chung Hsueh Yu Chi Sheng Chung 17:294-297(1999).
DR EMBL: AF062349; AAC72885.1; -.
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
SQ SEQUENCE 1704 AA; 193763 MW; 385526D0DA56FD1D CRC64;

Query Match 84.2%; Score 1915; DB 5; Length 1704;
Best Local Similarity 93.9%; Pred. No. 3.8e-85;
Matches 370; Conservative 8; Mismatches 10; Indels 6; Gaps 2;

QY 38 TDDDDKAMADIGSIEGRGTMAISVTMDNILSGFENEYDVYILKPLAGVYRSLLKKQIEKNI 97
```





```
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 539 AA; 61076 MW; 72186F1412C57D1C CRC64;

Query Match 51.8%; Score 1178; DB 5; Length 539;
Best Local Similarity 56.9%; Pred. No. 6.3e-50;
Matches 227; Conservative 58; Mismatches 90; Indels 24; Gaps 4;

Qy 34 PDLGTDKADAMIGSTEGRTMAISVTMDNILSGFENYDVLYLPLAGVYSLKKQI 93
Db 144 PFGESEED--YDDLQGVVTGEAVTPSV-IDNILSKIENEYEVLYLPLAGVYSLKKQL 200

Qy 94 EKNIFTNMLNDILNSRLKRRKYFLDVLESDLMQFKHISSNEYIEDSFKLLNSEQKNT 153
Db 201 ENNVMTNVNVKDIILNSRFNFKNVLESDLPYKDLTSSNYVVKDPYKFLNKEKRD 260

Qy 154 LLKSYKIKESVENDIKFAQEGISYIEKVLAKYKDDLESIKKVIKEEKEKFPSPPTTP 213
Db 261 FLSSYNIKDSIDTIDINFANDVLGYKILSEKYSDDLDSIKKYI----- 304

Qy 214 SPKATDEQKESKFLPFLTNIETLYNNLVNKIDYDLINLKAINDCNVEKDEAHVKITKL 273
Db 305 ---NDKOGENEKYLPLFNNIETLYKTVDKIDLVFVHLEAKVLNITYEKSNEVEKIKEL 360

Qy 274 SDLAIDDKIDLFPNPDYFEAIIKLLINDDTTKDMLGKLLSTGLV-QNFPNTIISKLIEGK 332
Db 361 NYLTIQDKLADFKNNFVGIADLSTDYNNHLLTKFLSTGMVFENLAKTVLSNLLDGN 420

Qy 333 FQDMLNLSOHCYKQCPENSGCFRHLDERECKCLLNKYGKQKCVENPNTCNEENGG 392
Db 421 LQGLMNLISQHCYKQCPENSGCFRHLDERECKCLLNKYGKQKCVENPNTCNEENGG 480

Qy 393 CDADATCTEEDSGSRKKITCECTKPDSPYPLFDGIFCSS 431
Db 481 CDADAKTEEDSGSGKKITCECTKPDSPYPLFDGIFCSS 519

RESULT 37
Q03999 PRELIMINARY; PRT; 400 AA.
AC Q03999; 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1 precursor (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RA Blackman M.J., Ling I.T., Nicholls S.C., Holder A.A.;
RT "Proteolytic processing of the Plasmodium falciparum merozoite surface protein-1 produces a membrane-bound fragment containing two epidermal growth factor-like domains.";
RL Submitted (MAY-1991) to the EMBL/GenBank/DBJ databases.
DR EMBL; M64681; AAA29709.1;
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00008; EGF; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Signal.
FT NON_TER 1 1
FT SIGNAL <1 24 POTENTIAL.
FT CHAIN 25 286 MEROZOITE SURFACE PROTEIN 1.
FT CHAIN 287 400 MEROZOITE SURFACE PROTEIN 1.
SQ SEQUENCE 400 AA; 45824 MW; 537F075058626AC2 CRC64;

Query Match 51.6%; Score 1175; DB 5; Length 400;
Best Local Similarity 56.6%; Pred. No. 6.4e-50;
Matches 226; Conservative 59; Mismatches 90; Indels +24; Gaps 4;

Qy 34 PDLGTDKADAMIGSTEGRTMAISVTMDNILSGFENYDVLYLPLAGVYSLKKQI 93
Db 5 PFGESEED--YDDLQGVVTGEAVTPSV-IDNILSKIENEYEVLYLPLAGVYSLKKQL 61

Qy 94 EKNIFTNMLNDILNSRLKRRKYFLDVLESDLMQFKHISSNEYIEDSFKLLNSEQKNT 153
Db 62 ENNVMTNVNVKDIILNSRFNFKNVLESDLPYKDLTSSNYVVKDPYKFLNKEKRD 121

Qy 154 LLKSYKIKESVENDIKFAQEGISYIEKVLAKYKDDLESIKKVIKEEKEKFPSPPTTP 213
Db 122 FLSSYNIKDSIDTIDINFANDVLGYKILSEKYSDDLDSIKKYI----- 165
```

```
QY 214 SPAKTDQKKESKFLPFLNIETLYNNLVNKIDYILNKLKAKINDCNVEKDEAHVKITKL 273
Db 166 ----NDKQGENEKYLPFLNIETLYKTVDNKIDLFVHLEAKVNLNTYKSNVEYKIKEL 221
QY 274 SOLKATDDKIDLFKNPYDFAEATKKLINDTKKMDLGLKLLSTGLV-QNFPNTIISKIEGK 332
Db 222 NYLKTITQDKLADFKKNNFVGIADLSTDYNNHNNLLTKFLSTGMVFENLAKTVLSNLLDGN 281
QY 333 FQMLNISQHCQVKKQCPNSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENNGG 392
Db 282 LOGMLNISQHCQVKKQCPNSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENNGG 341
QY 393 CDADATCTEEDSGSRKKITCTCKTPDPSYPLFDGIFCSS 431
Db 342 CDADAKCTEEDSGSGNGKKITCTCKTPDPSYPLFDGIFCSS 380

RESULT 38
Q25984
ID Q25984 PRELIMINARY; PRT; 539 AA.
AC Q25984;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TremBLrel. 20, Last annotation update)
DE Major merozoite surface protein (fragment).
MSPI.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSPI) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13356; BAA02617.1;
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00008; EGF; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 539 AA; 61115 MW; 3788015F3127CB9E CRC64;

Query Match 51.6%; Score 1175; DB 5; Length 539;
Best Local Similarity 56.6%; Pred. No. 8.8e-50;
Matches 226; Conservative 59; Mismatches 90; Indels 24; Gaps 4;

QY 34 PDLTGDDDDKAMADIGSIEGRGTMAISVTMDNITLSGFENEYDVIYKPLAGVYRSLLKQI 93
Db 144 PIFGESEED--YDDLQGVVTGEAVTPSV-IDNLSKIENEYEVLYLKPLAGVYRSLLKQI 200
QY 94 EKNIFTFNLNDILNSRLKRRKYFLDVLESOLMOPKHSSNEYIIEDSFKLNSQKNT 153
Db 201 ENNVMTFNVNVDILNSRKNREKNFKNVLESOLIPYKDLTSSNVYVVKDPYKFLNKEKRD 260
QY 154 LKSKYIKESVENDIKFAQEGISYVEKVLAKYKDDLESIKKVIKEEKEFPSSPTTP 213
Db 261 FLSSYNYIKSDIDTDFNANDVGLVYKILSEKYSKDSLSIKKI----- 304
QY 214 SPAKTDQKKESKFLPFLNIETLYNNLVNKIDYILNKLKAKINDCNVEKDEAHVKITKL 273
Db 305 ----NDKQGENEKYLPFLNIETLYKTVDNKIDLFVHLEAKVNLNTYKSNVEYKIKEL 360
QY 274 SOLKATDDKIDLFKNPYDFAEATKKLINDTKKMDLGLKLLSTGLV-QNFPNTIISKIEGK 332
Db 361 NYLKTITQDKLADFKKNNFVGIADLSTDYNNHNNLLTKFLSTGMVFENLAKTVLSNLLDGN 420
QY 333 FQMLNISQHCQVKKQCPNSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENNGG 392
Db 421 LOGMLNISQHCQVKKQCPNSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENNGG 480

RESULT 40
Q25984
ID Q25984 PRELIMINARY; PRT; 539 AA.
AC Q25984;
DT 01-MAY-2000 (TremBLrel. 13, Created)
```

```
QY 393 CDADATCTEEDSGSRKKITCTCKTPDPSYPLFDGIFCSS 431
Db 481 CDADAKCTEEDSGSGNGKKITCTCKTPDPSYPLFDGIFCSS 519

RESULT 39
Q25966
ID Q25966 PRELIMINARY; PRT; 539 AA.
AC Q25966;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TremBLrel. 20, Last annotation update)
DE Major merozoite surface protein (fragment).
MSPI.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSPI) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13357; BAA02618.1;
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00008; EGF; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 539 AA; 61145 MW; 7B7DE90C1D0ACDE7 CRC64;

Query Match 51.6%; Score 1175; DB 5; Length 539;
Best Local Similarity 56.6%; Pred. No. 8.8e-50;
Matches 226; Conservative 59; Mismatches 90; Indels 24; Gaps 4;

QY 34 PDLTGDDDDKAMADIGSIEGRGTMAISVTMDNITLSGFENEYDVIYKPLAGVYRSLLKQI 93
Db 144 PIFGESEED--YDDLQGVVTGEAVTPSV-IDNLSKIENEYEVLYLKPLAGVYRSLLKQI 200
QY 94 EKNIFTFNLNDILNSRLKRRKYFLDVLESOLMOPKHSSNEYIIEDSFKLNSQKNT 153
Db 201 ENNVMTFNVNVDILNSRKNREKNFKNVLESOLIPYKDLTSSNVYVVKDPYKFLNKEKRD 260
QY 154 LKSKYIKESVENDIKFAQEGISYVEKVLAKYKDDLESIKKVIKEEKEFPSSPTTP 213
Db 261 FLSSYNYIKSDIDTDFNANDVGLVYKILSEKYSKDSLSIKKI----- 304
QY 214 SPAKTDQKKESKFLPFLNIETLYNNLVNKIDYILNKLKAKINDCNVEKDEAHVKITKL 273
Db 305 ----NDKQGENEKYLPFLNIETLYKTVDNKIDLFVHLEAKVNLNTYKSNVEYKIKEL 360
QY 274 SOLKATDDKIDLFKNPYDFAEATKKLINDTKKMDLGLKLLSTGLV-QNFPNTIISKIEGK 332
Db 361 NYLKTITQDKLADFKKNNFVGIADLSTDYNNHNNLLTKFLSTGMVFENLAKTVLSNLLDGN 420
QY 333 FQMLNISQHCQVKKQCPNSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENNGG 392
Db 421 LOGMLNISQHCQVKKQCPNSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENNGG 480

RESULT 40
Q25966
ID Q25966 PRELIMINARY; PRT; 539 AA.
AC Q25966;
DT 01-MAY-2000 (TremBLrel. 13, Created)
```

01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
Major merozoite surface protein (Fragment).  
MSP1.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
SEQUENCE FROM N.A.  
RX MEDLINE=93295445; PubMed=8515786;  
RA Jongwutives S., Tanabe K., Kanbara H.;  
RT "Sequence conservation in the C-terminal part of the precursor to the  
major merozoite surface proteins (MSP1) of Plasmodium falciparum from  
field isolates."  
RL Mol. Biochem. Parasitol. 59:95-100(1993).  
DR EMBL; D13358; BAA02619.1; -;  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR001245; Tyr-kinase.  
DR Pfam; PF00008; EGF; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
KW EGF-like domain; Merozoite.  
FT NON\_TER 1  
SQ SEQUENCE 539 AA; 61075 MW; C70C2E100EC4A101 CRC64;  
  
Query Match 51.6%; Score 1175; DB 5; Length 539;  
Best Local Similarity 56.6%; Pred. No. 8.8e-50;  
Matches 226; Conservative 59; Mismatches 90; Indels 24; Gaps 4;  
  
QY 34 PDLGTDGDDKAMADIGSTEGRTWAISVTMDNIIILSGFENEYDVYIYLPAGVYSLKKOI 93  
DB 144 PIFGESED--YDGLGQVVGTEAVTPSV-IDNIIILSKTENEYEVLYLPAGVYSLKKQL 200  
  
QY 94 EKNIFTNLNLDILNSRLKRRKYFLDVLSDLMQFHHISSNEYIIEDSEKLLNSEQKNT 153  
DB 201 ENNVMTENVVVKDILNSRFNKNFNKLVESDLIPYKDLTSSNVVVKDPYKFLNKKRDK 260  
  
QY 154 LLKSYKIKESVENDIKFAQEGISYIEKVLAKYKDDLESIKKVIKEKEKPPSPPTTPP 213  
DB 261 FLSSYNYIKDSIDTDINFANDVLGYKILSEKYSLDLSIKKYI----- 304  
  
QY 214 SPAKTDEQKESKFLPFTNLTETLYNNLVNKKIDDYILNLRKAKINDCNVKEDEAHVKITKL 273  
DB 305 ----NDKQGENEKYLPFNLTETLYNKTVDKIDLVFVHLEAKVLYTYEKSNNVEVKIKEL 360  
  
QY 274 SDLKAIDDKIDLFKNPYDFEAIKKLINDTKKMLGKLLSTGLV-QNPPNTIISKLIIEGK 332  
DB 361 NYLKTIQDKLADFKNKNNFVGIADLSTDYNNHLLTKFLSTGMVFENLAKTVLSNLLDGN 420  
  
QY 333 FODMLNISQHCVKKQCPNSGCGFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGG 392  
DB 421 LQGLMLNISQHCVKKQCPNSGCGFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGG 480  
  
QY 393 CDADATCEDSGSRKKITCECTKPDSPPLFDGIFCSS 431  
DB 481 CDADAKTEEDSGSGNKKITCECTKPDSPPLFDGIFCSS 519

Search completed: May 12, 2003, 10:19:20  
Job time : 95 secs

GenCore version 5.1.4\_p5-4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 19, 2003, 14:37:01 ; Search time 2567 Seconds  
(without alignments)  
4886.369 Million cell updates/sec

Title: US-10-057-531A-2

Perfect score: 2275

Sequence: 1 MHHHHHSSGLVPRGSMKE.....TCECTKPDSPFLDGIKCSS 431

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlh  
-O=/cgn2.1/USPTO.spool/US10057531/runat\_12052003.091031\_21666/app\_query.fasta\_1.583  
-DB=GenEmbl -QFT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=40 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10057531@cgn.1.1.1687 -runat\_12052003.091031\_21666 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl: \*  
1: gb\_ba: \*  
2: gb\_htg: \*  
3: gb\_in: \*  
4: gb\_om: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sts: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vi: \*  
15: em\_ba: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_mu: \*  
20: em\_om: \*  
21: em\_or: \*  
22: em\_ov: \*  
23: em\_pat: \*  
24: em\_ph: \*  
25: em\_pl: \*  
26: em\_ro: \*  
27: em\_scs: \*  
28: em\_un: \*

29: em\_vi: \*  
30: em\_htg\_hum: \*  
31: em\_htg\_inv: \*  
32: em\_htg\_other: \*  
33: em\_htg\_mus: \*  
34: em\_htg\_pln: \*  
35: em\_htg\_rod: \*  
36: em\_htg\_mam: \*  
37: em\_htg\_vrt: \*  
38: em\_sy: \*  
39: em\_htgo\_hum: \*  
40: em\_htgo\_mus: \*  
41: em\_htgo\_other: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1970	86.6	1249	12	AF395808 Synthetic
2	1969.5	86.6	5312	3	PFEZSA1B
3	1965.5	86.4	1959	3	PFEZSA1C
4	1946.5	85.6	1729	3	PFEZSA14
5	1946.5	85.6	1729	3	PFEZSA20
6	1946.5	85.6	5392	3	PFE190G1
7	1946.5	85.6	5754	3	PFEZSA1A
8	1942	85.4	1119	3	PFEZSA1B
9	1937	85.1	1726	3	PFEZSA1C
10	1935.5	85.1	5541	3	PFEZSA14
11	1935	85.1	1119	3	PFEZSA20
12	1935	85.1	1956	3	PFEZSA1A
13	1933	85.0	1119	3	PFEZSA1B
14	1932.5	84.9	5276	3	PFEZSA1C
15	1931	84.9	1726	3	PFEZSA14
16	1931	84.9	1726	3	PFEZSA20
17	1931	84.9	1726	3	PFEZSA1A
18	1929	84.8	1119	3	PFEZSA1B
19	1928	84.7	1119	3	PFEZSA1C
20	1924	84.6	1119	3	PFEZSA14
21	1922.5	84.5	1116	3	PFEZSA20
22	1921	84.4	1726	3	PFEZSA1A
23	1919	84.4	1119	3	PFEZSA1B
24	1919	84.4	1119	3	PFEZSA1C
25	1917.5	84.3	1116	3	PFEZSA14
26	1917.5	84.3	1116	3	PFEZSA20
27	1917.5	84.3	1116	3	PFEZSA1A
28	1917	84.3	1956	3	PFEZSA1B
29	1915	84.2	1726	3	PFEZSA1C
30	1915	84.2	1726	3	PFEZSA14
31	1915	84.2	1726	3	PFEZSA20
32	1915	84.2	1726	3	PFEZSA1A
33	1915	84.2	1726	3	PFEZSA1B
34	1915	84.2	5243	3	PFEZSA1C
35	1915	84.2	5243	3	PFEZSA14
36	1915	84.2	5271	3	PFEZSA20
37	1906.5	83.8	1116	3	PFEZSA1A
38	1901.5	83.6	1116	3	PFEZSA1B
39	1900.5	83.5	1116	3	PFEZSA1C
40	1900	83.5	1726	3	PFEZSA14
41	1178	51.8	1636	3	PFEZSA20
42	1178	51.8	1636	3	PFEZSA1A
43	1175	51.6	1203	3	PFEZSA1B
44	1175	51.6	1636	3	PFEZSA1C
45	1175	51.6	1636	3	PFEZSA14

ALIGNMENTS

RESULT 1

AF395808 1249 bp mRNA linear SYN 20-AUG-2001  
 LOCUS Synthetic construct melittin signal peptide/MSPl-42 fusion protein  
 mRNA, complete cds.  
 ACCESSION AF395808  
 VERSION AF395808.1 GI:15213639  
 KEYWORDS synthetic construct.  
 SOURCE synthetic construct  
 ORGANISM artificial sequences.  
 REFERENCE 1 (bases 1 to 1249)  
 AUTHORS Pang,L.Y., Hashimoto,C.N., Tam,L.Q., Meng,Z.Q., Hui,S.N. and  
 Ho,K.K.  
 TITLE In vivo expression and immunological studies of the 42 kDa  
 carboxyl-terminal processing fragment of Plasmodium falciparum  
 merozoite surface protein 1 (MSPl-42) in the Baculovirus-silkworm  
 system  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 1249)  
 AUTHORS Pang,L.Y., Hashimoto,C.N., Tam,L.Q., Meng,Z.Q., Hui,S.N. and  
 Ho,K.K.  
 TITLE Direct Submission  
 JOURNAL Submitted (28-JUN-2001) Department of Biochemistry, The Chinese  
 University of Hong Kong, Shatin, Hong Kong  
 FEATURES  
 source  
 1..1249  
 /organism="synthetic construct"  
 /isolate="3D7"  
 /db\_xref="taxon:32630"  
 26..1228  
 /note="honeybee melittin signal sequence fused with 42 kDa  
 carboxyl-terminal processing fragment of Plasmodium  
 falciparum isolate 3D7 merozoite surface protein 1"  
 /codon\_start=1  
 /transl\_table=1  
 /product="melittin signal peptide/MSPl-42 fusion protein"  
 /protein\_id="AAK92098.1"  
 /db\_xref="GI:15213640"  
 /translation="MKFLVNVALFVMVVIYIYADPSPMAISVTMDNLSGFENEYD  
 VVLPGLAGYVRLSKQIEKNIFTNLNDIILNSLRKRYFLDVLSDLMQFKHIS  
 SNVYIEDFKLNSQKNTLLSKYIKESVENDIKFAQEGISYEVKLVAKYKDDLE  
 SIKKVIKEEKKEPSPPTPPSPAKTDEQKESKLPFLTNITLNNLVNKIDDL  
 INLKAKINDNVEKDEKHVITKLSLKAIDDKIDLFKNPYDFEAKKLINDDTKDM  
 LKLLSTGLVQNPNTIISKLIKFKQDMINIHOHQCKQCPENSGCFRHLDEREC  
 KCLNLYKQEGDKGVENPNPNPTNNNGCCDADATCTEDDSGSRKKIKTCETKPDVPL  
 FDGIKCS"  
 -BASE COUNT 518 a 180 c 185 g 366 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1.62e-129 Length: 1249  
 Score: 1970.00 Matches: 375  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 86.59% Indels: 0  
 DB: 12 Gaps: 0  
 US-10-057-531A-2 (1-431) x AF395808 (1-1249)  
 Qy 57 MetAlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspVal 76  
 |||||||  
 Db 101 ATGCAATATCTGTACAAATGGTAATATCTCTCAGGATTTGAAATATGATATGATGTT 160  
 |||||||  
 Qy 77 IleTyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsn 96  
 |||||||  
 Db 161 ATATATTAAACCTTAGCTGGAGTATATAGACGCTTAAAAACAAATTTGAAAAAAC 220  
 |||||||  
 Qy 97 IlePheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLys 116  
 |||||||  
 Db 221 ATTTTACATTTAAATTTGAACGATATCTTAAATTCACGCTCTTAAGAAACGAAAA 280  
 |||||||  
 Qy 117 TyrPheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGlu 136  
 |||||||

Db 281 TATTCTTAGATGATTAGAAATCTGATTAAATGCAATTTAAACATATATCTCAATGAA 340  
 |||||||  
 Qy 137 TyrIleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLys 156  
 |||||||  
 Db 341 TACATTATTGAAGATTTCATTTAAATTTATTGAATTCAGAACAAAAAACACACTTTTAAA 400  
 |||||||  
 Qy 157 SerTyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIle 176  
 |||||||  
 Db 401 AGTTACAAATATATAAAGAAATCAGTAGAAATGATATTAATTTGCACAGGAAGGTATA 460  
 |||||||  
 Qy 177 SerTyrTyrGluLysValLeuAlaLysTyrLysAspLeuGluSerIleLysLysVal 196  
 |||||||  
 Db 461 AGTTATTATGAAAGGTTTATAGCGAAATATAGGATGATTTAGATCAATTTAAAAAGTT 520  
 |||||||  
 Qy 197 IleLysGluGluLysGluLysPheProSerProProThrThrProProSerProAla 216  
 |||||||  
 Db 521 ATCAAGAAGAAAGGAGAGAGTTCCCATCATCACCAACCAACACCTCCCTGCCAGCA 580  
 |||||||  
 Qy 217 LysThrAspGluGlnLysLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThr 236  
 |||||||  
 Db 581 AAAACAGACGAAACAAAGAGAAAGTAAGTTCCTTCCATTTTAAACAAACATTGAGACC 640  
 |||||||  
 Qy 237 LeuTyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIle 256  
 |||||||  
 Db 641 TTATACAATAACTTAGTTAATAAATTAAGATTTACGATTACTTAATTAATTAAGCGCA 700  
 |||||||  
 Qy 257 AsnAspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeu 276  
 |||||||  
 Db 701 AACGATTGTATGTTGAAGAGATGAACCATGTTAAATTAACATACTTAGTGATTTA 760  
 |||||||  
 Qy 277 LysAlaIleAspLysLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLys 296  
 |||||||  
 Db 761 AAAGCAATTGATGACAAATAGATCTTTTAAAAACCCCTTACGACTTCGAAGCAATTA 820  
 |||||||  
 Qy 297 LysLeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeu 316  
 |||||||  
 Db 821 AAATTGATAATGATGATACGAAAAAGATATGCTTGCAAAATTTACTTAGTACAGGATTA 880  
 |||||||  
 Qy 317 ValGlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMet 336  
 |||||||  
 Db 881 GTTCAAAATTTTCTCTAATACATAATATCAAAATTAATTTGAAGGAAAAATTCGAAGATG 940  
 |||||||  
 Qy 337 LeuAsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPhe 356  
 |||||||  
 Db 941 TTAACATTTTCCACAAACCAATGCGTAAAAAACAATGTCAGAAAAATTTCTGGATGTTTC 1000  
 |||||||  
 Qy 357 ArgHisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluLysAsp 376  
 |||||||  
 Db 1001 AGACATTTAGATGAAGAGAGAAATGTAATGTTTATTAATTTACAAACAAGAGGTGAT 1060  
 |||||||  
 Qy 377 LysCysValGluAsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAsp 396  
 |||||||  
 Db 1061 AAATGCTGTTGAAATCCAAATCCTACTTGTAAACGAAAAATATGTTGGATGTCAGAT 1120  
 |||||||  
 Qy 397 AlaThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThr 416  
 |||||||  
 Db 1121 GCCACATGTACCGAAGAGATTCAGGTAGCAGCAGAGAAAAATCACATGTGAATGTACT 1180  
 |||||||  
 Qy 417 LysProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431  
 |||||||  
 Db 1181 AAACCTGATCTTATCCACTTTTCGATGGTATTTCTGCAGTTCC 1225  
 |||||||  
 RESULT 2  
 LOCUS PFMEZSA1B 5312 bp DNA linear INV 21-FEB-2002  
 DEFINITION P.falciparum gpl190 (MSAL, MSPL, PMMSA) for precursor of major  
 merozoite surface antigens.  
 ACCESSION Z35327  
 VERSION 235327.1 GI:929795  
 KEYWORDS gpl190 gene; merozoite surface antigen; MSAL gene; MSPL gene; PMMSA  
 gene.  
 SOURCE malaria parasite P. falciparum.  
 ORGANISM Plasmodium falciparum



Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

1 (bases 1 to 5312)

pan.W., Tolle.R., and Bujard.H.

A direct and rapid sequencing strategy for the Plasmodium

falciparum antigen gene gp190/MSA1

Mol. Biochem. Parasitol. 73 (1-2), 241-244 (1995)

96123395

8577332

2

Tolle.R., Bujard.H. and Cooper.J.A.

Plasmodium falciparum: variations within the C-terminal region of

merozoite surface antigen-1

Exp. Parasitol. 81 (1), 47-54 (1995)

95354793

7628566

3 (bases 1 to 5312)

Tolle.R.

Direct Submission

Submitted (18-JUL-1994) Ralf Tolle, Abt. Prof. H. Bujard, Zentrum

für Molekulare Biologie (ZMBH), Univ. Heidelberg, Im Neuenheimer

Feld 282, 69120 Heidelberg, Germany

On Aug 5, 1995 this sequence version replaced gi:535249.

Location/Qualifiers

1..5312

/organism="Plasmodium falciparum"

/isolate="NF54"

/db\_xref="taxon:5833"

/chromosome="9"

/clone="307A"

/dev\_stage="blood stage"

/gene="gp190, MSA1, MSP1, PMSA"

109..5271

/gene="gp190, MSA1, MSP1, PMSA"

109..5271

/codon\_start=1

/product="precursor of the major merozoite surface

antigens"

/protein\_id="CAA84556.1"

/db\_xref="GI:929796"

/db\_xref="SPTREMBL:Q25922"

/translation="MKIIFLCFLPFIITOCVTHSYQELVKKLEALEDAVLTCYS

LQKEMVNEIEITPKGASQASQASQASQASQASQASQASQASQASQASQASQ

SPSGTSPSRNNTLPRNTSSGASPPASDASDASDASDASDASDASDASDASD

ELFDNLNMLTCDNIHGPKYLDIGYEINELLYKLFYDLRLAKLNDVCANDYCOI

PNFLKRALNEDVLKLVFGYKPLDNIDNVGMEDYIKNNKTITANTNELIEGSKK

TDQKNADNEEGKKLYQADYLDSTYKQLEAHNLI SVLEKRIIDLTKNNENIKLL

DKINEIKNPANGSNPTNLLDKNKIEIEHEEKIKETAKTKFNIDSLTDPLELEY

YLREKNKYVDVTPKSDPTKSVQIPKVPYNGIVYPLPLTDHNSLAADNDKNSYGL

MPHPTKEIKELIITDNKERKIFINNIKKIDLEKNINHTKEONKKLLEDYEKSKD

YELLEKFTENAFNNFDKVDVDFISARTYTNVEKQRYNNKFSNNNSYNNVQKLKK

ANGLEDYSLRGISSEKFNHYITLTGTLEADIKLIEIKSENKILEKNFKFLHS

ALGSEVDIVLQVQVLLIKIEDLRKIELFKNAQLKDSIHVPNIYKPKQNKPEPY

YLIVLKKVDKLEKFPKVKMDLKEQAVLSSTQPLVAASETTEGDSHTHTLSQSG

ETVETETETETVCHTHTVTITLPPTOPSPKPKVVENSTEHKSDNSQALTKTV

YLKLDLEFTKSVICHYILVSNSSMDKLLLEVNTPEENELKSCDPLDLLFNQIN

NIPAMKSLDSMNNDQLHFFELYKEMIIYHLKKEENHKKLLEEQKITGTSST

SPGNTVTAQSAHNSQNSQNSNTQNGVAVSSGPVAVVEESHDPLTVLISND

LGKIVSLNLGNKTKPNPLITSTTEMERFENILKNNDYTDNDIKQVKSNSVIT

GLTQKNAINDIKKIDTLQSLFDLYNKYKLDLRLFNKKELQDMQKQIKLL

KEOLSKLNNHNVNLFQNSVFENKKEAEIETENTLENTKILLKHKVGLVKN

GESSPLKTLSEVSIQEDYANLEKERVLSKIDGLNDNLHLGKKLSFLSSGLHLI

TELKVEYKNTGTPSPENNNKVNALKSYENFLPEAKVTIVTPPQDVTPTSPUSV

RVSGSGSKKEETIPTSGSLTLEQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV

QVVTGSAIVMDNISLGFENEYDVIYLPKLAGVYRSKQKQIEKNITFTNMLNDILN

SLRKKRYLDVLESLOMQFKHSSNEYIIEDSFKLLNSBKNTLLKSYKIKESVEN

DIKFAQEGSYXPKVLAKYKDDLESIKVKIKBEKEKFPSPPTTPSPKTDQKES

KFLPFTNITLYLNNLVNKIDDIYLLNKKAINDCNVKEKDEAHVKIKLSLDLMDKI

DLFPKPYDEAIKLLNDITKDKMLGLSTGLVQNPFTIISKLIEGKQDMILISQ

HQCKKQCPENSCFRHLDRECKCLNKKQEGDKCVENPNPTCNENNGCCDADTC

TEEDSGSRKKITCECTKPDSPYPLFDGIFCCSSNFLGISFLILMLILYSFI"

109..165

/gene="gp190, MSA1, MSP1, PMSA"

2328 a 703 c 705 g 1576 t

sig\_peptide

BASE COUNT

# ORIGIN

## Alignment Scores:

Prod. No.: 8.01e-129

Score: 1969.50

Percent Similarity: 97.46%

Best Local Similarity: 95.69%

Query Match: 86.57%

Indels: 5

Gaps: 1

DB: 3

US-10-057-531A-2 (1-431) x PFMEZSALB (1-5312)

Qy 38 ThrAspAspAspLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMet 57

Db 4042 TCCGAGATAATGACGAATATTTAGATCAAGTAGTAACCTGCAGAA----- 4086

Qy 58 AlaIleSerValThrMetAspAsnIleIleuSerGlyPheGluAsnGluTyrAspValIle 77

Db 4087 GCAATATCTGTCACAATGGGATAATATCTCTCAGGATTTGAAATGAATATGATGTTATA 4146

Qy 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97

Db 4147 TATTTAAACCTTTAGCTGGAGTATATAGAAAGCTTAAAAAACAATTTGAAAAACAATT 4206

Qy 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgGlyTyr 117

Db 4207 TTTACATTTAAATTTAAATTTGACGATATCTTAAATTTACGCTCTTAAGAAACGAAATAT 4266

Qy 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137

Db 4267 TTCTTAGATGATTAGAAATCTGATTTAATGCAATTTTAAACATATATCTCAATGAATAC 4326

Qy 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157

Db 4327 ATTATTGAAGATTCATTTAAATTTGCAATTTTAAACATATATCTCAATGAATAC 4386

Qy 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177

Db 4387 TACAAATATATAAAGAAATCAGTAGAAATATGATTTAAATTTGCACAGGAGGTATAAGT 4446

Qy 178 TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysValIle 197

Db 4447 TATTATGAAAGGTTTGTAGCGAAATATAGGATGATTAGTAATCAATTTAAAAAGTTATC 4506

Qy 198 LysGluGluLysGluLysPheProSerSerProProThrThrProProSerProAlaLys 217

Db 4507 AAAGAAGAAAAAGGAGAGTTCCTCCATCATCACCAACACACCTCCGTCACCAAGCAAA 4566

Qy 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237

Db 4567 ACAGACGAACAAAAAGAGGAAAGTAAGTTCCTTCATTTTAAACAAACATTGAGACCTTA 4626

Qy 238 TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn 257

Db 4627 TACAATAACTTAGTTAATAAATTCAGATTACTTAATTAATTAAGGCAACAAATTAAC 4686

Qy 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277

Db 4687 GATTGTAATGTTGAAAAAGATGAAGCACATCTTAAAAATACTAACTTAGTATTAAAAA 4746

Qy 278 AlaIleAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys 297

Db 4747 GCAATTGATGACAAAAATAGATCTTTTAAAAACCCCTTACGACTTTTCAAGACAAATTTAAAAA 4806

Qy 298 LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal 317

Db 4807 TTGATAATGATGATACGAAAAAGATGCTTGCATAATTTACTTAGTACAGGATTAGTT 4866

Qy 318 GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeu 337

Db 4867 CAAAAATTTCTCTAATAACAATAATCAAAATTAATTTGAAGGAAAAATTTCAAGATATGTTA 4926

Qy 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357

```

Db 4927 ACATTTTCACACACCAATGCGTAAACAAACATGTCAGAAATTTCTGGATGTTTCAGA 4986
QY 358 HIsLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377
Db 4987 CATTTAGATGAAGACGAATGTAATCTTATTAAATTAACAACAAGAGTGATAA 5046
QY 378 CysValGluAsnProAsnProThrCysAsnGluAsnGlnGlyCysAspAlaAspAla 397
Db 5047 TCTGTTGAAATCCCAATCCTACTTGTAAACGAAATAATGTTGGATGTCATGCAGATGCC 5106
QY 398 ThrCysThrGluGluAspSerGlySerArgLysLysLysLysLysLysLysLysLys 417
Db 5107 ACATGTACCGAAGAGATTCAGTACGACAGAAAGAAATCACATGTGAATGTACTAAA 5166
QY 418 ProAspSerTyrProLeuPheAspGlyLysPheCysSerSer 431
Db 5167 CTGATCTTATCCACTTTTCAGTGGTATTTCTGCAGTTC 5208

RESULT 3
PFMEZSAIC 1959 bp DNA linear INV 30-OCT-1995
P.falciiparum (NF7) gene for merozoite surface antigen 1.
P.35328 M19144
P.35328.1 GI:535251
merozoite surface antigen 1 gene.
Plasmodium falciparum.
Plasmodium falciparum.
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
Peterson,M.G., Coppel,R.L., McIntyre,P., Langford,C.J., Woodrow,G.,
Brown,G.V., Anders,R.F. and Kemp,D.J.
Variation in the precursor to the major merozoite surface antigens
of Plasmodium falciparum
Mol. Biochem. Parasitol. 27 (2-3), 291-301 (1988)
88142999
2449612
2 (bases 1 to 1959)
Tolles.R.
Direct Submission
Submitted (18-JUL-1994) Ralf Tolles, Abt. Prof. H. Bujard, Zentrum
fur Molekulare Biologie, (ZMBH), Univ. Heidelberg, Im Neuenheimer
Feld 282, Heidelberg, 69120, Germany
3 (bases 1 to 1959)
Tolles.R., Bujard,H. and Cooper,J.A.
Plasmodium falciparum: variations within the C-terminal region of
merozoite surface antigen-1
Exp. Parasitol. 81 (1), 47-54 (1995)
95354793
7628566

FEATURES
source
Location/Qualifiers
1..1959
/organism="Plasmodium falciparum"
/isolate="FCH5/NF7"
/db_xref="taxon:5833"
/chromosome="9"
/dev_stage="blood stage"
1..1959
/gene="gp190, MSA1, MSP1, PMMSA"
<1..1959
/gene="gp190, MSA1, MSP1, PMMSA"
/codon_start=1
/product="merozoite surface antigen 1 gene, 3+-end"
/protein_id="CAA84557.1"
/db_xref="GI:535252"
/db_xref="SPTREMBL:Q25923"
/translaton="FNKKKELGQDKMOIKLTLKLEKSLNPNHVNLFNSVF
FNKKKEIAETENTLTKLLKHYKGLVYNGESSPLKTISEVQIEDNYANLE
KFRVLSKGLNDNLHGKKKLSFLSSGLHLITLKEVIKNNKNTGSPSSNNKV
NEALKSYENLPEAKVTVTVPQDTPSPSVYRSGSGSGSTKEETQIPTSGSLTE
LQQVQSQDEEDSLVLPFIPEGSENDDEYLDQVVTGRIASVTMDNLSGFENEYD
VYIKPLAGVYRSLLKQIEKNIFTNLNLDNLSRLKRYKFLDVLDESILMOFKHIS
SNEVIIESFKLNSEQKNLLSKYIKESVENDIKFAQEGISYYEKLAKYKDDLE

```

```

SIKKVKEEKEFPSPPTTPSPAKTDEQKESKFLPFLTNITLYNNLVNKIDYDL
INLAKINDCNVEKPDIAHVKITKLSLDKIDDKIDLFKNPKYDFEAKILNDTKKDM
LGKLLSTGLVONFNTIISKLTIEGFODMLNISOHCVKKOCPEKSGFRLHDEREC
KCLLNYKQEGDKCVENPNPTCNENNGCGDADATCTEKDSGSRKKITCECTKPDYPL
PDGIFCSSNFGLISFLILMLILYSFI"
BASE COUNT 855 a 266 c 573 t
ORIGIN

```

## Alignment Scores:

```

Pred. No.: 5, 37e-129 Length: 1959
Score: 1965.50 Matches: 376
Percent Similarity: 97.46% Conservative: 8
Best Local Similarity: 95.43% Mismatches: 5
Query Match: 86.40% Indels: 5
DB: 3 Gaps: 1

```

US-10-057-531a-2 (1-431) x PFMEZSAIC (1-1959)

```

QY 38 ThrAspAspAspLysAlaMetAlaAspLleGlySerIleGluGlyArgGlyThrMet 57
Db 730 TCCGAAGATAATGACGAATATTTAGATCAAGTAGTAAGTGGAGAA----- 774
QY 58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77
Db 775 GCAATATCTGTCAATGGATAATATCTCTCAGGATTTGAAATGAATATGATGTTATA 834
QY 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97
Db 835 TATTTAAACCTTTAGCTGGAGTATATAGAGCTTAAACAAATTTGAAAAACATTT 894
QY 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117
Db 895 TTTACATTTAAATTTAAATTTGAACGATATCTTAAATTCACGCTTTAAGAAACGAAATAT 954
QY 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137
Db 955 TTTCTAGATGATTAGAAATCTGATTTAATGCAATTTAAACATATATATCTCAATGAATAC 1014
QY 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157
Db 1015 ATTATTGAAGATTCATTTAAATTTATTTGAATTCAGAACAAAAACACACTTTTAAAGT 1074
QY 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177
Db 1075 TACAAATATATAAAAGATTCAGTAAAGATGATATTAATTTTGCACAGAGGTATACT 1134
QY 178 TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValIle 197
Db 1135 TATTATGAAAGGTTTACCGAATATATAGGATGATTTAGAAATCAATTTAAACAAAGTTATC 1194
QY 198 LysGluGluLysGluLysPheProSerSerProProThrThrProProSerProAlaLys 217
Db 1195 AAAGAAGAAAAAGGAGAGTTCCTCATCATCACCAACCAACACCTCCGTCACCACAAAA 1254
QY 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237
Db 1255 ACAGACGAACAAAAAGAGAAAGTAAAGTTCCCTTCCATTTTAAACAAACATTTGAGACCTTA 1314
QY 238 TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn 257
Db 1315 TACAATACTTAGTTAATAAATTTGACCATTTACTTTAATTAATTAACGCAAGAAAGTTAAC 1374
QY 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277
Db 1375 GATTGTAATGTTGAAAAAGATGAAGACACATGTTAAATTAACATACTTAACTTAGTATTAAAA 1434
QY 278 AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys 297
Db 1435 GCAATTGATGACAAATAGATCTTTTAAAAACCCCTTACGACTTTCGAGACCAATTAAGAAA 1494
QY 298 LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal 317
Db 1495 TTGATAAATGATGATACGCAAAAAAGATATGCTTGGCAAAATTACTTAGTACAGGATTAGTT 1554

```

QY 318 GlnAspPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeu 337  
 DB 1555 CAAATTTTCCCTAATAACATAATATCAATAATTAATTGAAGGAAATTCACAGATATGTA 1614  
 QY 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357  
 DB 1615 AACATTTCACACACCAATCGCTTAAACAAACAAATGTCGCAAAATTCCTGGATGTTTCA 1674  
 QY 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377  
 DB 1675 CATTTAGATGAAGAAGAAGTAATGTTTATTAAATACAAACAAAGAGTGATAA 1734  
 QY 378 CysValGluAsnProAsnProThrCysAsnGluAsnGlyGlyCysAspAlaAspAla 397  
 DB 1735 TGTGTTGAAATCCCAATCCTACTTGTAAACGAAATTAATGGTGGATGATGCAGATGCC 1794  
 QY 398 ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417  
 DB 1795 ACATGTACCGAAAGAAATTCAGGTAGCAGCAGAAAGAAATCAGATGTGAATGTACTAAA 1854  
 QY 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431  
 DB 1855 CCTGATCTTATCCACTTTTTCGATGATATTTCTGCAGTTC 1896

## RESULT 4

PFAMSP814

LOCUS

DEFINITION

P. falciparum DNA for the precursor to the major merozoite surface

proteins, C-terminal.

DI3345

DI3345.1

GI:391790

EGF-like domains; major merozoite surface protein precursor.

KEYWORDS

Plasmodium falciparum (isolate 814) merozoite, DNA.

SOURCE

Plasmodium falciparum

ORGANISM

Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.

REFERENCE

Jongwutives, S., Tanabe, K. and Kanbara, H.

Amino acid sequence conservation in the C-terminal part of the precursor to

the major merozoite surface proteins (MSP1) of Plasmodium

falciparum from field isolates

Mol. Biochem. Parasitol. 59 (1), 95-100 (1993)

93295445

2 (bases 1 to 1729)

Jongwutives, S.

Direct Submission

Submitted (05-OCT-1992) Sanchai Jongwutives, Institute of Tropical

Medicine, Nagasaki Univ., Dept. of Protozoology, 12-4

Sakamoto-machi, Nagasaki 852, Japan (tel:0958-47-2111(ex.3747),

Fax:0958-47-6607)

Submitted (05-OCT-1992) to DDBJ by:

Sanchai Jongwutives

Department of Protozoology

Institute of Tropical Medicine

Nagasaki University

12-4 Sakamoto-machi

Nagasaki 852

Japan

Phone: 0958-47-2111x3747

Fax: 0958-47-6607.

Location/Qualifiers

1. .1729

/organism="Plasmodium falciparum"

/db\_xref="taxon:5833"

/chromosome="9"

/dev\_stage="merozoite"

1. .1715

/gene="MSP1"

1. .1715

/gene="MSP1"

/codon\_start=3

/product="major merozoite surface protein precursor"

/protein\_id="BAA02606.1"

CDS

## FEATURES

source

gene

CDS

/db\_xref="GI:391791"

/translaltion="PLKTLSEVSIQIEDNYANLEKFRVLKIDGKLNLDNLHLGKKKLS  
 FLSSGLHLITELKWKVKNKNTGSPSENNKKNVAKLSYENFLPEAKYTVTPPFG  
 PDVTPSPLSVRVSGSGSTKEETQPTSGSLLETLQVQVQSONYDEEDSLVLPFG  
 ESEDNDYLDQVVTGEAISVTMDNLSGFENYDVYILKPLAGVYRSUKQTEKNIT  
 FNLNLDILNSRLKRRKFLDVLDSLMQFKHISSENYIIESFKLNSQKTEKLLK  
 YKIKESVENDIKFAQEGISYKVLAKYKDDLESIKKVIKEKFKFPSPPTPPSP  
 AKTDQKKESKFLPFLTNIETLYNLNKKIDYILNKKAKINDCAVKEKDEAHVKTKL  
 SPDKALDDKIDLFKNTDPEATKKLINDTCKMDLGLKLLSTGLVQNPNTIISKLEIG  
 KQDMLNISQHCYVKQCPENSCFRHLDERECKLLNYKQEGKCKVENPNTCNEN  
 NGCCDADATCTEEDSGSRKKITCETCPDPSYPLDFGIFCSSNLFGLSLMLLIL  
 YSFI"

BASE COUNT 742 a 234 c 246 g 507 t

ORIGIN

## Alignment Scores:

Score: 1,01e-127 Length: 1729  
 Pred. No.: 1946.50 Matches: 374  
 Percent Similarity: 96.70% Conservative: 7  
 Best Local Similarity: 94.92% Mismatches: 8  
 Query Match: 85.56% Indels: 5  
 DB: 3 Gaps: 1

US-10-057-531A-2 (1-431) x PFAMSP814 (1-1729)

QY 38 ThrAspAspAspLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMet 57  
 DB 486 TCCGAAGATAATGACCAATATTTAGATCAAGTAGTAGTACGAGAA----- 530  
 QY 58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77  
 DB 531 GCAATATCTGTCAATGGAATATATCTCTCAGGATTTGAAATGAATATGATGTATA 590  
 QY 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97  
 DB 591 TATTTAAACCTTTAGCTGGAGTATATAGAGCTTAAAAACCAATTTGAAAAACATT 650  
 QY 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117  
 DB 651 ATTACATTTAAATTTAAATTTGAACGATATCTTAAATTCACGCTTAAAGAAACGAAATAT 710  
 QY 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137  
 DB 711 TCTTAGATGTATTAGAATCTGATTTAATGCAATTTAAACATATATCTCTCAAAATGAATAC 770  
 QY 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLysSer 157  
 DB 771 ATTATTGAAGATTCATTTAAATTTATTTGAATTCAGACACAAAAACACACCTTTTAAAAAGT 830  
 QY 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177  
 DB 831 TACAATATATATAAAGAATCAGTAGAAAAATGATATAATTTGCACAGGAAGGTATAGT 890  
 QY 178 TyrTyrGluLysValLeuAlaLysTyrLysAspLeuGluSerIleLysLysValIle 197  
 DB 891 TATTATCAAAAGGTTTTAGCGAAATATAAGGATGATTAGATCAATTAATAAAGTTATC 950  
 QY 198 LysGluGluLysGluLysPheProSerSerProThrThrProProSerProAlaLys 217  
 DB 951 AAAGAAGAAAGAGAGAGTTCCCATCATCATCACCAACACACCTCGCTCACCAGCAAAA 1010  
 QY 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237  
 DB 1011 ACAGACGACAAAGAAAGAGAGTAAAGTTCCTTCCATTTTAAACAAACATTGAGACCTTA 1070  
 QY 238 TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysAsn 257  
 DB 1071 TACAATAACTTAGTTAATAAAATTGACGATTACTTAATTAACTTAAAGCGCAAGATTAAC 1130  
 QY 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277  
 DB 1131 GATTGTAGTGTGAAAGAGATGAAGCACATGTTAAAAATACTAACTTAGTGTATTTAAAA 1190

Qy 278 AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys 297  
 Db 1191 GCAATTGATGACAAATAGATCTTTTAAAAACACTAACGACTTCGAAGCAATTAATAAAA 1250  
 Qy 298 LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal 317  
 Db 1251 TTGATAAATGATGATACGAAAAAGATGCTTGGCAAAATTTACTTAGTACAGGATTAGTT 1310  
 Qy 318 GlnAsnPheProAsnThrIleIleSerLysLeuIleGlyLysPheGlnAspMetLeu 337  
 Db 1311 CAAATTTTCTTANTACATAATATCAAAATTAATGAAGGAAATTCGAAGATATGTTA 1370  
 Qy 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357  
 Db 1371 AACATTTTCCACAAACCAATGCGTAAAAAACAATGCTCCAGAAAAATTCGGATGTTTCA 1430  
 Qy 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377  
 Db 1431 CATTAGATGAAGAGAGAATGTAATGTTTATTAAATTAACAAGAAAGGTTGATAAA 1490  
 Qy 378 CysValGluAsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAla 397  
 Db 1491 TGTGTGAAATCCAAATCCTTACTTGTAAACGAAAAATAATGTTGATGTCAGATGCC 1550  
 Qy 398 ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417  
 Db 1551 ACATGTACCGAAGAAGATTTCAGGTAGCAGCAGAAAGAAATTCACATGTGAATGACTAAA 1610  
 Qy 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431  
 Db 1611 CCTGATCTTATCCACTTTTCGATGGTATTTTCGACAGTTCC 1652

RESULT 5  
 PFAMSPM20  
 LOCUS

DEFINITION P. falciparum DNA for the precursor to the major merozoite surface proteins, C-terminal.  
 ACCESSION D13343  
 VERSION D13343.1 GI:391822  
 KEYWORDS EGF-like domains; major merozoite surface protein precursor.  
 SOURCE Plasmodium falciparum (isolate MAD20) merozoite, DNA.  
 ORGANISM Plasmodium falciparum  
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

REFERENCE 1  
 Tanabe,K., Mackay,M., Goman,M. and Scaife,J.G.  
 Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmodium falciparum  
 J. Mol. Biol. 195 (2), 273-287 (1987)  
 88011243

JOURNAL  
 MEDLINE  
 REFERENCE 2 (bases 1 to 1729)  
 Jongwutives,S., Tanabe,K. and Kanbara,H..  
 Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP1) of Plasmodium falciparum from field isolates  
 Mol. Biochem. Parasitol. 59 (1), 95-100 (1993)  
 93295445

JOURNAL  
 MEDLINE  
 REFERENCE 3 (bases 1 to 1729)  
 Jongwutives,S.  
 Direct Submission  
 Submitted (05-OCT-1992) Somchai Jongwutives, Institute of Tropical Medicine, Nagasaki Univ., Dept. of Protozoology: 12-4  
 Sakamoto-machi, Nagasaki 852, Japan (Tel:0958-47-2111(ex.3747), Fax:0958-47-6607)  
 Submitted (05-OCT-1992) to DDBJ by:  
 Somchai Jongwutives  
 Department of Protozoology  
 Institute of Tropical Medicine  
 Nagasaki University  
 12-4 Sakamoto-machi  
 Nagasaki 852  
 Japan  
 Phone: 0958-47-2111x3747  
 Fax: 0958-47-6607.

COMMENT

FEATURES  
sourceLocation/Qualifiers  
1. .1729

/organism="Plasmodium falciparum"

/db\_xref="taxon:5833"

/chromosome="9"

/dev\_stage="merozoite"

1. .1715

/gene="MSP1"

&lt;1. .1715

/gene="MSP1"

/codon\_start=3

/product="major merozoite surface protein precursor"

/protein\_id="BAA02604.1"

/db\_xref="GI:391823"

/translation="PLKLTSEVSIQTEDNYANLEKFRALSKIDGKLNLDNLHLGKKLLS  
 FLSSGLHLITELKEVTKNNKNTGNSPENNKKVNEALKSVENFLPEAKVTVTVTPQ  
 PDVTPSLVRVSGSGSKEETQIPTSGSLLTELQOVVOLQNYDEDDSLVLPFG  
 ESDNDEYLDQVVTGSAISVDMNIIISGFENEDYVILKPLAGYVYSLKQIKENLIT  
 FNLNDILNSRLKKRYFLDVLSDLMQPKHSSNEYIIEEDSKLNSQKNTLLKS  
 YKIKESVENDIKFAQEGISYKVLAKYKDDYILNKLKINDCNVKEKDPSPPTPSP  
 AKTDEQKESKFLPFTNITLNNLVNKIDDLILNKLKINDCNVKEKDPSPPTPSP  
 SDLKAIDDKIDLPKNTDPEAIKKLINDPTKDKMLGKLGKLTGLVQNFPTNISKLEIG  
 KFDMLNIHQOCVKKOCPSGCFRHLDERECKLLNKKQSGDKCVENPNTCAEN  
 KGCDDATCTEEDSGSSKKITCECTKPSYPLFDGIFCNSNFIQISFLILMLIL  
 YSF"

BASE COUNT 742 a 234 c 246 g 507 t

## ORIGIN

## Alignment Scores:

Pred. No.: 1,01e-127 Length: 1729  
 Score: 1946.50 Matches: 374  
 Percent Similarity: 96.70% Conservative: 7  
 Best Local Similarity: 94.92% Mismatches: 8  
 Query Match: 85.56% Indels: 5  
 DB: 3 Gaps: 1

US-10-057-531A-2 (1-431) x PFAMSPM20 (1-1729)

Qy 38 ThrAspAspAspLysAlaMetAlaAspIleGlySerIleGlyArgGlyThrMet 57

Db 486 TCCGAAGATAATGACGAATATTAGATCAAGTAGTAACTGGAGAA----- 530

Qy 58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValle 77

Db 531 GCAATATCTGTCACAATGGATATATCTCTCAGGATTTGAAATGAATATGATGTTATA 590

Qy 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97

Db 591 TATTTAAACCTTTAGCTGGAGTATATAGAGCTTAAAAAACAAATTTGAAAAACATT 650

Qy 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117

Db 651 ATTACATTTAATTTAAATTTGAACGATATCTTAAATTCACGCTCTTAAGAAACGAAATAT 710

Qy 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137

Db 711 TTCTTAGATGATTAGAAATCTGATTAAATGCAATTTAAACATATATCTCAATGAATAC 770

Qy 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157

Db 771 ATTATGACGATTCATTTAAATTTGAACGATATCTTAAATTCACGCTCTTAAAGAAAGT 830

Qy 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177

Db 831 TACAATATATAAAGAAATCAGTAGAAATGATATAATTTAAATTTGCACAGGAGGTATAAGT 890

Qy 178 TyrTyrGluLysValLeuAlaLysTyrLysAspLeuGluSerIleLysLysValle 197

Db 891 TATTATGAAAGGTTTGTAGCGAAATATAAGGATGATTAGAAATCAATTAATAAGTTATC 950

Qy 198 LysGluGluLysGluLysPheProSerSerProThrThrProProSerProAlaLys 217

Db 951 AAAGAAGAAAGGAGAGAGTTCCTCATCATCACCAACACACCTCCCTCACCAGCAAA 1010





NOQSNASTNTONGVAVSSGPAVVEESHDPITVLISNDLKGIVSLNLGNKTKVPNP  
 LTISTTEKEFYENILKNNDYFNDDIKQFVKSNSKITGLTETQKNALNDEIRKLKD  
 TLOLSFDLYNKYKLRFLPNKRELQGDQKQIKRLTLKLEQLESKLNLNPNHVLQ  
 NFSVFNKKKAETAEKTLNKLKLVKGLVGYNGESSPLKTLSEVSIOTEDN  
 YANLEKFRALSKIDGKNDNLHLGKKLSFLSSGLHLHLITELKEVKKNNKTGNSPSE  
 NKKKFKALSKIDGKNDNLHLGKKLSFLSSGLHLHLITELKEVKKNNKTGNSPSE  
 SLLELQOVQVQYDEDDSLVLPFGESEDNDEYLDQVVTGEATSVTNDNLISGF  
 ENEDVYILKPLAGYRSLKQIEKNIITFNLDNLNLSRLKRRKFLDVLDESLLMQ  
 FKHSSNIEYIPEDSFKLLNSQKNTLLKSYKIKESVENDIKFAQEGISYVEKVLAKY  
 KDDLESIKVYIKEKEKEFPSPPTPPSPAKTDQOKKESKFLPTLTNLETLYNNLVNK  
 IDVYLINLKAINDCNVEKDEAHVYKTKLSDLKALDDKIDFLKNTNDEFAKLKLNDD  
 TKDKMLGLLSTGLVONFPNTIISKLIEGKFQDMNLSIHOQCVRKQCPENSGCFPHLD  
 ERECKCLNLTQKQKDCVKNPNPTCNENGGCDADATCTEEDSGSSRKKITCBECKP  
 DSYPLDFGICFSSNFLGISFLILMLILYSFI  
 121..177  
 sig\_peptide  
 mat\_peptide

/gene="PMMSA"  
 178..5223  
 /product="major merozoite surface antigen"  
 BASE COUNT 2539 a 709 c 708 g 1798 t  
 ORIGIN 218 bp upstream of HindIII site.

Alignment Scores:  
 Pred. No.: 3,56e-127 Length: 5754  
 Score: 1946.50 Matches: 374  
 Percent Similarity: 96.70% Conservative: 7  
 Best Local Similarity: 94.92% Mismatches: 8  
 Query Match: 85.56% Indels: 5  
 DB: 3 Gaps: 1

US-10-057-531A-2 (1-431) x PFAMMSAA (1-5754)

QY 38 ThrAspAspAspAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMet 57  
 DB 3997 TCGAAGATAATGACGAATATTAGTCAAGTAGTACCTGGAGAA-----4041  
 QY 58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77  
 DB 4042 GCAATATCTGTACATATGATATATCTCTCAGGATTTGAAATGAATATGATGTATA 4101  
 QY 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysGlnIleGluLysAsnIle 97  
 DB 4102 TATTTAAACCTTTAGCTCGAGTATATAGAAGCTTAAACAAACAAATTCGAAACAAAT 4161  
 QY 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysTyr 117  
 DB 4162 ATTACATTTTAAATTTGAACGATATCTTAAATTCAGCTTAAAGAACGAAATAT 4221  
 QY 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137  
 DB 4222 TTCTTAGATGTATTAGAATCTGATTTAATGCAATTTAAACATATATCTCAATGAATAC 4281  
 QY 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLysSer 157  
 DB 4282 ATTATTGAAGATTCATTTAAATATTGAATTCAGAACAAACAAACACCTTTTAAAGAT 4341  
 QY 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177  
 DB 4342 TACAAATATATAAACAATCAGTAGAAATGATATTAAATTTGCACAGGAAGTATAAGT 4401  
 QY 178 TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysValIle 197  
 DB 4402 TATTATGAAAGGTTTTAGCGAAATATAAGGATGATTAGAATCAATTTAAAGATTTATC 4461  
 QY 198 LysGluLysLysLysPheProSerSerProThrThrProProSerProAlaLys 217  
 DB 4462 AAGAAGAAAGAGAGAGTTCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 4521  
 QY 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237  
 DB 4522 ACAGACGACAAAGAAAGAGTTCATCATCATCATCATCATCATCATCATCATCATCATCAT 4581  
 QY 238 TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn 257

DB 4582 TACAATAACTTGTAGTTAAATAAATTCAGGATTAATTAATTAAGCAAGATTAAC 4641  
 QY 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277  
 DB 4642 GATTGTAAATGTTGAAAAGATGAAGCACATGTTAAAATACTAAATCTTAGTATTTAAA 4701  
 QY 278 AlaIleAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys 297  
 DB 4702 GCATTTGATGACAAATAGATCTTTTAAAACACTAACGACTTCCAGCAATTAATAA 4761  
 QY 298 LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal 317  
 DB 4762 TTGATAAATGATGATACGAAAAAGATATGCTTGGCAAAATTAATTAGTACAGGATTA 4821  
 QY 318 GlnAsnPheProAsnThrIleLeuSerLysLeuIleGluLysPheGlnAspMetLeu 337  
 DB 4822 CAAATTTTCTTAATAACAATATATCAAAATTAATTTGAAGGAAATTTCCAAGATATG 4881  
 QY 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357  
 DB 4882 AACATTTACACACCAATCGGTAAACAAACATGTCAGAAATTTCTGGATGTTTCAGA 4941  
 QY 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377  
 DB 4942 CATTTAGATGAAAGAGAGAATGTAATGTTTATTAAATTACAAACAAGAGGTGATAA 5001  
 QY 378 CysValGluAsnProAsnProThrCysAsnGluAsnGlyCysAspAlaAspAla 397  
 DB 5002 TGTGTTGAAAATCCCAATCTCTTGTAAACGAAATTAATGGTGATGTCAGATGCC 5061  
 QY 398 ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417  
 DB 5062 ACATGTACCGAAGAGATTCAGTAGCAGCAGAGAAACAAATCATGTGATCTACTATA 5121  
 QY 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431  
 DB 5122 CCTGATTTCTTCCACTTTTCGATGATGATTTCTGCAATTC 5163  
 RESULT 8  
 PFU02728 1119 bp DNA linear INV 14-FEB-1995  
 Plasmodium falciparum clone K47A merozoite surface protein 1  
 (MSP-1) gene, 42 kDa C-terminal region, partial cds.  
 U20728  
 U20728.1 GI:664857  
 malarial parasite.  
 Plasmodium falciparum  
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 Shi,Y.-P., Alpers,M.P., Povoa,M.M., Nahlen,B.L., Oloo,A.G. and  
 Lal,A.A.  
 Sequence of the C-terminal region of merozoite surface protein-1  
 (MSP-1) in field-derived Plasmodium falciparum  
 Unpublished  
 Shi,Y.-P.  
 Direct Submission  
 Submitted (07-FEB-1995) Ya-Ping Shi, DpD/CDC 4770, Molecular  
 Vaccine Section, Bldg 22, Rm. 8F12, Buford Highway, Chamblee, GA,  
 30341 USA  
 Location/Qualifiers  
 1. .1119  
 /organism="Plasmodium falciparum"  
 /isolate="infected human blood sample from Papua New  
 Guinea"  
 /db\_xref="taxon:5833"  
 /clone="K47A"  
 1. .1119  
 /gene="MSP-1"  
 <1..>1119  
 /gene="MSP-1"  
 source  
 gene  
 CDS



```
/codon_start=1
/product="merozoite surface protein 1, 42 kDa C-terminal
/region"
/protein_id="AAA62219.1"
/db_xref="GI:664858"
/translation="AISTVMDNISLGFENEDVILYKPLACVYSLKKQIEKNIFTN
LNLNDILNSRLKKRYFLDVLDESDFKLNSENEYIEDSFKLNSEQKNTLKKSYK
YIKSEVNDIKFAQGGISYEKVLAKYKDDLESIKKVIKEERKFPSPPTPPSPAK
TDEQKESKFLPFTNLETETLYNNLVNIDYLYNLKAKINDCNVDEARVHTKLSL
LKATDDKIDLFKNPYDEAKKLINDDTKKDMLGKLLSTGLVQNPNTIISKLEGRF
QDMLNIHQHCVKQSCQPNQSCFRHDEBECKLLNYKQEDKCVENPNPTCNENNG
GCDADAKTEEDSNGKIKITCECTKDSYPLFDGIFCS"
BASE COUNT      485 a  147 c  161 g  326 t
ORIGIN
Alignment Scores:
Pred. No.:      1,32e-127      Length:      1119
Score:          1942.00      Matches:      369
Percent Similarity: 99.46%      Conservative: 2
Best Local Similarity: 98.93%      Mismatches: 2
Query Match:      85.36%      Indels:      0
Db:               3          Gaps:      0

US-10-057-531A-2 (1-431) x PFU20728 (1-1119)

Qy 58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77
Db 1 GCAATATCTGTCACAAATGGATAATATCTCTCAGGATTGAAATATGATGTTATA 60

Qy 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97
Db 61 TATTAAACCTTTAGCTGGAGTATATAGAGCTTAAAAAACAATTTGAAAAACAAT 120

Qy 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117
Db 121 TTTACATTAATTTAAATTTGAACGATATCTTAAATTCACGCTCTTAAGAAACGAAATAT 180

Qy 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137
Db 191 TTCCTAGATGATTAGAACTGATTAAATGCAATTTAAACATATATCTCAATCAATAC 240

Qy 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157
Db 241 ATTATTGAAGATTTCATTTAAATTTATTGAAATTCAGACACAAAAACACACTTTTAAAGT 300

Qy 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177
Db 301 TACAAATATATAAAGAAATCAGTAGAAATGATATTTAAATTTGCACAGGAGGTATAAGT 360

Qy 178 TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValIle 197
Db 361 TATTATGAAAAGGTTTATAGCGAAATATAAGGATGATTAGAAATCAATTTAAAAAGTTATC 420

Qy 198 LysGluGluLysGluLysPheProSerSerProProThrThrProProSerProAlaLys 217
Db 421 AAAGAAAGAAAGGAGAGAGTTCCCATCATCATCACCAACACACACCTCCGTCACACGAAAA 480

Qy 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237
Db 481 ACAGACGACAAAAAGAGAAAGTAAAGTTCCTTCATTTTAAACAAACATTGACACCTTA 540

Qy 238 TyrAsnAsnLeuValAsnLysIleAspAspTyrIleuIleAsnLeuLysAlaLysIleAsn 257
Db 541 TACAAATACCTAGTTAATAAATTTAGCATTACTTAAATTAACCTTAAAGGCAAGATTAAC 600

Qy 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277
Db 601 GATTGTAATGTTGAAAAGATGAAGCAGCATCTTAAATACTAACTAGTATGATTTAAAA 660

Qy 278 AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys 297
Db 661 GCAATTGATGACAAAATAGATCTTTTAAAAACCCCTTACGACTTCGAAGCAATTAATAA 720
```

```
Qy 298 LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal 317
Db 721 TTGATAAATGATGATACGAAAAAAGATATGCTTGGCAATTAATTACTTAGTACAGATTAGT 780

Qy 318 GlnAsnPheProAsnThrIleIleSerLysLeuIleGluLysPheGlnAspMetLeu 337
Db 781 CAAATTTTCCTTAATACAAATATATCAAAATTAATTAAGGAAAAATTCACAGATATGTTA 840

Qy 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357
Db 841 AACATTTCCACACACCAATGCGTAAAAAACAATGTCACAAAATTCGGATGTTTTCAGA 900

Qy 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377
Db 901 CATTTAGATGAAGAGAAGATGTAATGTTTATTAAATTTACAAACAAGAAGTGTATATAA 960

Qy 378 CysValGluAsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAla 397
Db 961 TGTGTTGAAAATCCAAATCCTACTTGTAAACGAAAAATATGTTGGATGTGATCGAGATGCC 1020

Qy 398 ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417
Db 1021 AAATGTACCGAAGAAGATTTCAGTAGCAACGGAAGAAATAACACATGTGAATGTACTATA 1080

Qy 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSer 430
Db 1081 CCGTATCTTATCCACTTTTCGATGGTATTTTCTGCAGT 1119

RESULT 9
PFAMSP841
LOCUS
DEFINITION P. falciparum DNA for the precursor to the major merozoite surface
proteins, C-terminal.
ACCESSION D13351
VERSION D13351.1 GI:391810
KEYWORDS EGF-like domains; major merozoite surface protein precursor.
SOURCE Plasmodium falciparum (isolate 841) merozoite, DNA.
ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 1726)
AUTHORS Jongwutiwes, S., Tanabe, K. and Kanbara, H.
TITLE Sequence conservation in the C-terminal part of the precursor to
the major merozoite surface proteins (MSP) of Plasmodium
falciparum from field isolates
JOURNAL Mol. Biochem. Parasitol. 59 (1), 95-100 (1993)
MEDLINE 93295445
REFERENCE 2 (bases 1 to 1726)
AUTHORS Jongwutiwes, S.
TITLE Direct Submission
JOURNAL Submitted (05-OCT-1992) Somchai Jongwutiwes, Institute of Tropical
Medicine, Nagasaki Univ., Dept. of Protozoology, 12-4
Sakamoto-machi, Nagasaki 852, Japan (Tel:0958-47-2111(ex.3747),
Fax:0958-47-6607)
COMMENT Submitted (05-OCT-1992) to DBJ by:
Somchai Jongwutiwes
Department of Protozoology
Institute of Tropical Medicine
Nagasaki University
12-4 Sakamoto-machi
Nagasaki 852
Japan
Phone: 0958-47-2111x3747
Fax: 0958-47-6607.
FEATURES
source
1. .1726
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/chromosome="9"
/dev_stage="merozoite"
1. .1712
/gene="MSP1"
<1. .1712
/gene="MSP1"
gene
CDS
```



Qv 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277

CDS.



(MSP-1) in field-derived Plasmodium falciparum  
Unpublished  
2 (bases 1 to 1119)  
Shi, Y.-P.  
Direct Submission  
Submitted (07-FEB-1995) Ya-Ping Shi, DPD/CDC 4770, Molecular  
vaccine Section, Bldg 22, Rm. 8F12, Buford Highway, Chamblee, GA,  
30341 USA

Location/Qualifiers  
1. 1119  
/organism="Plasmodium falciparum"  
/isolate="infected human blood sample from Papua New  
Guinea"  
/db\_xref="taxon:5833"  
/clone="K46B"  
1. 1119  
/gene="MSP-1"  
/gene="MSP-1"  
/codon\_start=1  
/product="merozoite surface protein 1, 42 kDa C-terminal  
region"  
/protein\_id="AA62218.1"  
/db\_xref="GI:664856"  
/translation="AISVTMDNIIISGFENEDYVILKPLAGVYSLRKKQIEKNFTFN  
LNLDILNRLKRRKYFLDVLDMQFKHSSNEYIIFEDSKLINSQKNTLLKSYK  
YIKESVNDIKFAOEGISYVEKVLAKYKDDLESIKVKEKEKFPSSPPTPPSPAK  
TDEQKESKFPFTNLETLYNNLVNKIDHYLINLKAINDCNVEKDHVKIKLSD  
LKAIDDKIDFKNPYDEAIIKLLINDTKDMGLKLLSTGLVQNPFTIISKLEGRF  
QMLNLSHQCVKQCPONSGFRHLDEREKKLLNYKQEGDKCVENPNPTCMENN  
GCDADAKTDEDSGSKKTKCTCKDPSYPLFDGIFCS"

BASE COUNT 485 a 148 c 160 g 326 t  
ORIGIN

Alignment Scores:  
Pred. No.: 4,09e-127 Length: 1119  
Score: 1935.00 Matches: 368  
Percent Similarity: 99.20% Conservative: 2  
Best Local Similarity: 98.66% Mismatches: 3  
Query Match: 85.05% Indels: 0  
DB: 3 Gaps: 0

US-10-057-531A-2 (1-431) x PFU20727 (1-1119)

Qy 58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77  
|||||  
Db 1 GCANATCTCTCACAATGATAATATCTCTCAGGATTTGAANAATGAATATGATGTTATA 60  
|||||

Qy 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97  
|||||  
Db 61 TATTTAAAACCTTTAGCTGGAGTATAGAAAGCTTAAAAAACAATTCGAAAAACATT 120  
|||||

Qy 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117  
|||||  
Db 121 TTTACATTTAAATTTAAATTTGAACGATATCTTTAAATTCACGCTTAAAGAAACGAAATAT 180  
|||||

Qy 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137  
|||||  
Db 181 TTCCTAGATGATATAGAAATCTGATTTAAACATATATATCTTCAATGAATAC 240  
|||||

Qy 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLysSer 157  
|||||  
Db 241 ATTATTGAAGATTCATTAAATTTATTGAATTCAGAACAAAAAACACACTTTTAAAAAGT 300  
|||||

Qy 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluLysIleSer 177  
|||||  
Db 301 TACAAATATATAAAGAATCAGTAGAAAATGATATTAATTTGCACAGGAAGGTATAAGT 360  
|||||

Qy 178 TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValIle 197  
|||||  
Db 361 TATATGAAAGGTTTTACGGAATATAAGGATGATTTAGATCAATTAATAAAGTTATC 420  
|||||

Qy 198 LysGluGluLysPheProSerProProThrThrProProSerProAlaLys 217  
|||||

JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source

Db 421 AAAGAAGAAAGGAGAAGTTCCCATCATCACCAACCAACACCTCCGTCACGCAAAA 480  
|||||

Qy 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237  
|||||

Db 481 ACAGACGAAACAAAGAAAGAAAGTAAGTTCCCTTCCATTTTAAACAACATTCAGACCTTA 540  
|||||

Qy 238 TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn 257  
|||||

Db 541 TACATAACTTAGTTAATAAATTTGACCATTACTTAATTAACTTAAAGCAAGATTAAC 600  
|||||

Qy 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277  
|||||

Db 601 GATTGTGAATGTTGAAAGACATGAAGCACATGTTAAATAACTTAACTTAGTGTGATTTAAAA 660  
|||||

Qy 278 AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaLysLys 297  
|||||

Db 661 GCAATTTGATGACAAAATAGATCTTTTAAAAACCCCTTACGACTTCGAAGCAATTAATAAAA 720  
|||||

Qy 298 LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal 317  
|||||

Db 721 TTGATAAATGATGATACGAAAAAGATATGCTTGGCAATTTACTTACTACAGGATTTAGTT 780  
|||||

Qy 318 GlnAsnPheProAsnThrIleLeuSerLysLeuIleGluGlyLysPheGlnAspMetLeu 337  
|||||

Db 781 CAAAATTTCTTAATACAAATAATCAAAATTAATTTGAAGAAAAATTCCAAGATATGTTA 840  
|||||

Qy 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357  
|||||

Db 841 ACATTTTCAACACACCAATGCGCTTAAAAACAATGTTCCAAAAATTTCTGGATGTTTCACA 900  
|||||

Qy 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377  
|||||

Db 901 CATTAGATGAAGAGAGAAGTAATGTTTATTAAATTACAAACAAGAGGTGATAA 960  
|||||

Qy 378 CysValGluAsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAla 397  
|||||

Db 961 TGTGTTGAAATCCCAATCTTACTTGTAAACGAAAAATTAATGTTGATGATGATGATGCC 1020  
|||||

Qy 398 ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417  
|||||

Db 1021 AAATGTACCGAAGAGATTTCAGTACGACGGAAGAAATACATGTGAATGATCTATAA 1080  
|||||

Qy 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSer 430  
|||||

Db 1081 CCTGATTTCTATCCACTTTTCGATGCTATTTCTGCGAGT 1119  
|||||

RESULT 12  
PFMEZSA1A  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source

PFMEZSA1A 1956 bp DNA linear INV 22-SBP-1995  
P.falciparum (RO-33) gene for merozoite surface antigen 1.  
235326 M35727  
235326.1 GI:535247  
merozoite surface antigen 1 gene.  
Plasmodium falciparum.  
Plasmodium falciparum.  
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
Tolte, R., Bujard, H. and Cooper, J.A.  
Plasmodium falciparum: variations within the C-terminal region of  
merozoite surface antigen-1  
Exp. Parasitol. 81 (1), 47-54 (1995)  
95354793  
7628566  
2 (bases 1 to 1956)  
Tolte, R.  
Direct Submission  
Submitted (18-JUL-1994) Ralf Tolte, Abt. Prof. H. Bujard, Zentrum  
fur Molekulare Biologie (ZMBH), Univ. Heidelberg, Im Neuenheimer  
Feld 282, Heidelberg, 69120, Germany  
Location/Qualifiers  
1. 1956  
/organism="Plasmodium falciparum"

```
/isolate="RO-33"
/db_xref="taxon:5833"
/chromosome="g"
/dev_stage="blood stage"
1..1956
/gene="gpi90, MSAL, MSP1, PMMSA"
<1..1956
/gene="gpi90, MSAL, MSP1, PMMSA"
/codon_start=1
/product="merozoite surface antigen 1 gene, 3+-end"
/protein_id="CA8455.1"
/db_xref="GI:535248"
/db_xref="SWISS-PROT:P19598"
/translocation="FNKKELGQDKMOIKLTLKLEQESKLSLNPNHNLQNSVF
FNKKAEAEAEENTLENTKILKHYKGLVYNGESSPLKTLSEVISIOTEDNYANLE
KPVLSKIDGLNDLHGLKKKSLSSGLHLITELKEVKNKNTGNSPSENKKV
NEALKSYENFLPAKVTVTVPQDPVTPSLSVRSVSGSGSPKEETOIPTSGSLITE
LQOVVOLQNYDEEDDSLVLPPIGESEDNDEYLDQVVTGEAISVMDNLSGSENYD
VIYKPLAGVYRSKLOKIEKNIFNLNLDILNSLKKRKYKFLDVEKSLQMKFHS
SNEYIIEDESKLNSOKNTLKSXYIKIESVENDIKFAQEGISYERVLAKYKDDLE
SIRKVIKESEKFPSPPTPPSPAKDEQKESKFLPFLNITLNLVNLVNTKDDYLI
NLKAKINDCNVEKDEAHVKITKLSLDKAIIDDKIDLPKNPYDFEAIKLLNDPTKQML
GKLLSTGLVONPPNTIISKLEGFQDMNLNISOHCQVKKOCFONSCGFRLHLDRECK
CLLNYKQEGDKVENPNTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDYPLF
DGIFCSNPLGLSFLILMLILYSPI"
```

BASE COUNT 853 a 264 c 265 g 574 t.

ORIGIN

Alignment Scores:

Pred. No.:	7 35e-127	Length:	1956
Score:	1935.00	Matches:	372
Percent Similarity:	96.70%	Conservative:	9
Best Local Similarity:	94.42%	Mismatches:	7
Query Match:	85.05%	Indels:	6
DB:	3	Gaps:	2

US-10-057-531A-2 (1-431) x PFMEZSALA (1-1956)

```
Qy 38 ThrAspAspAspLysAlaMetAlaAspIleGlySerIleGluGlyA-gGlyThrMet 57
Db 730 TCCCAAGATAATGACGAATATTAGATCAAGTAGTAGTGGAGAA----- 774
Qy 58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValile 77
Db 775 GCAATATCTGTCAACATGGGATAATATCTCTCAGGATTTGAAATGAATATGATGTTATA 834
Qy 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnile 97
Db 835 TATTTAAACCTTTAGCTGGAGTATATAGAGCTTAAAAACAAATTTGAAAAACATT 894
Qy 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117
Db 895 TTTACATTTAATTTAAATTTGACGATATCTTAAATTCAGCTCTTAAGAAACGAAATAT 954
Qy 118 PheLeuAspValLeuGluSerAspLeuMetClnPheLysHisIleSerSerAsnGluTyr 137
Db 955 TCTTAGATGATTAGAAATCTGATTTAATGCAATTTAAACATATATCTCAATGAATAC 1014
Qy 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157
Db 1015 ATTATTCAAGATTCATTTAAATATTGCAATTCAGACAAAAAACACACATTTAAAAAGT 1074
Qy 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177
Db 1075 TACAAATATATAAAGAAATCAGTAGAAATCATATTAATTTGCACAGGAGGTATAGT 1134
Qy 178 TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValile 197
Db 1135 TATTATGAAAGGTTTGTAGCGAAATATAAGATGATTAGATCAATTAATAAAGTTATC 1194
Qy 198 LysGluGluLysGluLysPheProSerSerProProThrThrProProSerProAlaLys 217
Db 1195 AAAGAAGAAAAAGGAG---TTCCCATCATCATCCACCAACACACCTCGCTCACCAGCAAA 1251
```

```
Qy 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrIleu 237
Db 1252 ACAGACCAACAAAGAGAGAAAGTAAGTCTCTCCATTTTAAACAAACATGTGACCTTA 1311
Qy 238 TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn 257
Db 1312 TACAATAACTAGTTAATAAAATTGACGATTACTTAATTAACTTAAAGCAAGCAATTAAAC 1371
Qy 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277
Db 1372 GATTGTAATGTTGAAAAGATGAAGCACATGTTAAATAAATACTAAGTATGATTTAAAA 1431
Qy 278 AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys 297
Db 1432 GCAATTGATGACAAAATAGATCTTTTAAACCCCTTACGACTTCGAAGCAATTAATAAA 1491
Qy 298 LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal 317
Db 1492 TTGATAATGATGATACGAAAAAGATATGCTTGGCAATTTACTTAGTACAGGATTAGTT 1551
Qy 318 GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeu 337
Db 1552 CAAAATTTTCTATATACAAATAATCAAAATTAATTGAAGAAAAATTCCAAGATATGTTA 1611
Qy 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357
Db 1612 AACATTTCAACAACACCAATGCGTAAAAAAACAATGTCCACAAAATTTCTGGATGTTT 1671
Qy 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377
Db 1672 CATTTAGATGAAAGAGAGAAATGTAATGTTTAAATTTACAAACAAAGAGGTGATAAA 1731
Qy 378 CysValGluAsnProAsnProThrCysAsnGluAsnGlnGlyCysAspAlaAspAla 397
Db 1732 TGTGTTGAAAATCCAAATCTCTTGTAAAGAAAATTAATGTTGATGTCAGATGCC 1791
Qy 398 ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417
Db 1792 AAATGTACCGAAGAAAGATTTCAGGTAGCAAGGAAAGAAAATTCACATGTGAATGTACT 1851
Qy 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431
Db 1852 CCTGATTTCTTATCCATTTTTCGATGGTATTTCCTGCGAGTTCC 1893
RESULT 13
PFU20732 1119 bp DNA linear INV 14-FEB-1995
LOCUS Plasmodium falciparum clone B17A merozoite surface protein 1
DEFINITION (MSP-1) gene, 42 kDa C-terminal region, partial cds.
ACCESSION U20732
VERSION U20732.1 GI:664865
KEYWORDS malarial parasite.
SOURCE Plasmodium falciparum
ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 1119)
AUTHORS Shi,Y.-P., Alpers,M.P., Povoa,M.M., Nahlen,B.L., Oloo,A.G. and Lal,A.A.
TITLE Sequence of the C-terminal region of merozoite surface protein-1 (MSP-1) in field-derived Plasmodium falciparum
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1119)
AUTHORS Shi,Y.-P.
TITLE Direct Submission
JOURNAL Submitted (07-FEB-1995) Ya-Ping Shi, DPD/CDC 4770, Molecular Vaccine Section, Bldg 22, Rm. 8F12, Buford Highway, Chamblee, GA, 30341 USA
FEATURES
source
1..1119
/organism="Plasmodium falciparum"
/isolate="infected human blood sample from Papua New Guinea"
```

Db	601	GATTGTAATCTGTGAAAAAATGAAGACACATGTTAAAAATACTAAACTTAGTCATTAAAA	660
Qy	278	AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys	297
Db	661	GCAATTGATCACAAAATAGATCTTTTAAAAACACTAACGACTTCGAGCAATTTAAAAA	720
Qy	298	LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal	317
Db	721	TTGATAAATGATGATACGAAAAAGATATGCTTGGCAAATTACTTAGTACAGGATTAGT	780
Qy	318	GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeu	337
Db	781	CAAAATTTTCCTAATACAATAATCAAAATTAATTGAAGGAAATTCAGAGATGTTA	840
Qy	338	AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPhearg	357
Db	841	AACATTTCAACAACCAATCGCTPAAAAACAATGTCCAGAAATTCGTGGATGTTTCAGA	900
Qy	358	HisLeuAspGluArgGluGluCysLysCysLysLeuLeuAsnTyrLysGlnGluGlyAspLys	377
Db	901	CATTTAGATCAAGAGAGAAGAAATGTAATGTTTATAATTACAAACAAGAGGTGATAAA	960
Qy	378	CysValGluAsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAla	397
Db	961	TGTGTTGAAAATCCAAATCCTACTTGTACAGAAAATAATGTTGGATGTGATGCAGATGCC	1020
Qy	398	ThrCystThrGluGluAspSerGlySerSerArgLysLysIleThrCysGluCysThrLys	417
Db	1021	ACATGTACCGAAGACAGATTCAGGTACACACAGACAGAAANATCCATGTGAATGTACTAAA	1080
Qy	418	ProAspSerTyrProLeuPheAspGlyIlePheCysSer	430
Db	1081	CCGTGATCTTATCCACTTTTCGATGATTTTCTGCAGT	1119

[illegible]

```

NSVYVQK LKKALSYLEDYSLRGISEKOPNHYITLKTGLEADIKKLTBEIKSENK
ILEKRGKLTSHANASLEVYDVLQVQKLLIKKIEDRLKIELFKNAQKDSIHVP
NIYKPNQKPEPYLVILKKEVDKLEFIPKVKMLKKEQAVISSITOPLVAASETTID
GGHSTHTLSQSGETEETETETETHTTITVITLPPKVKVSVSEIEHKSNDNSQ
ALTVTVLKKIDFELTKSYICHKYLIVSNSSMDOKLLEVYNLTPEEENELKSCDPLD
LFNTONIPAMSYLSDSNNDLOHLFPELYOKEMYLYLKLKEENHKKLLERQKQIT
GTSSTSPGNTVTNAGASHTNSQNSQNASSTNTQNGVAVSSGPVAVESHDPITV
LSISNDKGVLSLNLGNKTKVPNPLTISTEMBEKFYENILKNNDTIFNDDIKQFVK
NSKVITGLTETQKALNDEIKKLDITLQLSFDLYNKKLDLRLFNKKKELGQDKMQI
KKLLKLEQESKLSNNPHNVQNFQSVFNNKKKEAEIAETENTLENTKLLKHYKG
LVKYNGESSPLKTLSEVSIOTEDYANLEKFRVLSKIDGKLDNKLHLGKKKLSFLSS
GLHOLITJELKVIKNKNTGNSPENKKVNEALKSYENELPEAKVTVTVTPPOPDVT
PSPLSRVSGSGSTKETQIPTSGSLTELLQVQVQLQNDDEDDSLVLPFPGESD
NDEYLDQVQGEALSVTMDNLSGFENEYDVIYIKPAGVYRSKLRKQIEKNIFTNL
LDNLNLRKKRYFLVDLESDLMQFHHISNEYIIEEDSKLNSKQNTLLKSKYKI
KESVENDIKFAEQESLYEYKVLAKYKDDLESIKKVIKEKEKPSPTTPPSPAKTD
EOKKESFIPPLTNIETLYNNLVNKIDDLILNKAINDCNVEKDEAHVKTLLSDLK
AIDDKIDLKFNHNDFAKLLINDTKMDKLLSLTGLVONFPNTILSKIEKQFD
MLNLSQKCVKQKCPENSGCFRHDRECKLNLKQEGDKCVENPNPCNENNGGC
DADAKTEEDSGSGKKITCETKPDSPYLPFDGIFCSSNPLGISFILMLILYSFI

BASE COUNT      2325 a      686 c      711 g      1554 t
ORIGIN

Alignment Scores:
Pred. No.:      3,11e-126      Length:      5276
Score:          1932.50      Matches:      371
Percent Similarity: 96.45%      Conservative: 9
Best Local Similarity: 94.16%      Mismatches: 9
Query Match:      84.95%      Indels:      5
DB:              3          Gaps:      1

US-10-057-531a-2 (1-431) x PFAGP195 (1-5276)

Qy 38 ThrAspAspAspAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMet 57
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3952 TCCGAAGATAATGACGAATATTAGATCAAGTAGTAACGGAGNA----- 3996

Qy 58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValle 77
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3997 GCAATATCTGTCACAAATGGAATATCTCTCAGGATTGTGAATAATGAATGATGTTATA 4056

Qy 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4057 TATTTAAACCTTTAGCTGGAGTATATAGACCTTAAAAAACAATTTGAAAAACAAT 4116

Qy 98 PheThrPheAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4117 TTTACATTTAATTTAAATTTGAACGATATCTTAATTCACGCTTAAAGAACGAATAT 4176

Qy 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4177 TTCTTAGATGATTAGATCTGATTAAATGCAATTTAAACATATATATCTCAATCAATAC 4236

Qy 138 IleIleGluAspSerPheLysLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4237 ATTATTGAAGATTCAATTTAAATTTAGATTGCAACAAAAAACAACACTTTTAAAAAGT 4296

Qy 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4297 TACAATATATAAAGATCAGTACAAATGATTTAATTTGACAGGAGGTATAGT 4356

Qy 178 TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysValle 197
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4357 TATTATGAAAAGGTTTTAGCGAAAATATAGGATGATTTAGAAATCAATTTAAAAAGTTATC 4416

Qy 198 LysGluGluLysGluLysPheProSerProThrThrProThrProSerProAlaLys 217
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4417 AAAGAAGAAAAGGAGAGTCCCATCATCACCAACCAACACCTCCCGTCACGACAAAA 4476

Qy 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4477 ACAGACGACAAAGAAAGAAAGTAAGTTCCTTCATTTTAAACAAACATTTGAGACCTTA 4536
```

```

Qy 238 TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn 257
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4537 TACAATACTTAGTTTAATAAAAAATTGACGATTACTTAATTAATTAAGCAAGATTAAC 4596

Qy 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4597 GATTGTAATGTTGAAAAAGATGAAGACATGTTTAAATAAATACTTAACTTAGTATTAAAA 4656

Qy 278 AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaLysLys 297
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4657 GCAATTGATGACAAATAATAGATCTTTTAAAAACCAATACGACTTCGATGCAATTTAAAA 4716

Qy 298 LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal 317
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4717 TTGATAAATGATGATACGAAAAAAGATATGCTTGGCAAAATTACTTAGTACAGGATTAGTT 4776

Qy 318 GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeu 337
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4777 CAAATTTTCTTAATACAAATAATATCAAAATTAATTTGAAGGAAAAATTCAGATATGTTA 4836

Qy 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4837 AACATTTCCACACCAACCAATGCGTAAAAAACAATGTCAGAAAAATTCGGATGTTTCAGA 4896

Qy 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4897 CATTTAGATGAAAGAGAGAATGTAATGTTTATTAATTAATTAACAAAGAGTGCATAAA 4956

Qy 378 CysValGluAsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAla 397
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4957 TGTGTTGAAATCCAAATCCTACTTGTAAACGAAAAATAATGTTGGATGTGATGTCAGATGCC 5016

Qy 398 ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 5017 AAATGTCGGAAGAGATTTCAGGTAGCAACGGAAGAAAAATCACATGTGAATGTACTTAA 5076

Qy 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 5077 CCGTATCTTATCCACTTTTTCGATGTTATTTTCTGCAGTTCC 5118

RESULT 15
PFAMSP8221
LOCUS      1726 bp      DNA      linear      INV 29-MAY-2002
DEFINITION P. falciparum DNA for the precursor to the major merozoite surface
            proteins, C-terminal.
ACCESSION  D13347
VERSION    D13347.1 GI:391794
KEYWORDS   EGF-like domains; major merozoite surface protein precursor.
SOURCE     Plasmodium falciparum (isolate 822/1) merozoite, DNA.
ORGANISM   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE  1 (bases 1 to 1726)
AUTHORS    Jongwutives S., Tanabe K. and Kanbara H.
TITLE      Sequence conservation in the C-terminal part of the precursor to
            the major merozoite surface proteins (MSP1) of Plasmodium
            falciparum from field isolates
JOURNAL    Mol. Biochem. Parasitol. 59 (1), 95-100 (1993)
MEDLINE    93295445
REFERENCE  2 (bases 1 to 1726)
AUTHORS    Jongwutives S.
TITLE      Direct Submission
JOURNAL    Submitted (05-OCT-1992) Somchai Jongwutives, Institute of Tropical
            Medicine, Nagasaki Univ., Dept. of Protozoology; 12-4
            Sakamoto-machi, Nagasaki 852, Japan (Tel:0958-47-2111(ex.3747),
            Fax:0958-47-6607)
COMMENT    Submitted (05-OCT-1992) to DDBJ by:
            Somchai Jongwutives
            Department of Protozoology
            Institute of Tropical Medicine
            Nagasaki University
            12-4 Sakamoto-machi
            Nagasaki 852
```





## COMMENT

Submitted (05-OCT-1992) to DDBJ by:  
Somchai Jongwutiwes  
Department of Protozoology  
Institute of Tropical Medicine  
Nagasaki University  
12-4 Sakamoto-machi  
Nagasaki 852  
Japan  
Phone: 0958-47-2111x3747  
Fax: 0958-47-6607

## FEATURES

source  
Location/Qualifiers

1..1726

/organism="Plasmodium falciparum"

/db\_xref="taxon:5833"

/chromosome="9"

/dev\_stage="merozoite"

1..1712

/gene="MSPI"

<1..1712

/gene="MSPI"

/codon\_start=3

/product="major merozoite surface protein precursor"

/protein\_id="BAA02614.1"

/db\_xref="GI:391815"

/translation="PLKTLSEVSIQEDNTYANLEKFRALSKIDGKLNNDLHLGKKILS  
FLSGHLHLTELEKEVINKNYTGNSPSENKKVNEALKSYENFFPEAKVTVVTPFQ  
PDVTPSLVRVSGSGSTKEETQIPTSGSLTELQVQVQVQVDEDDSVTVVTPFQ  
ESEDNDVTVGTGEAISVTMDNILSGFENEDVYIKPLAGVYRSLKKQIEKNIT  
FNLNLDILNSRLKKRYFLVDLESILQMFKHSSNEYIITEDSKILNSPKQNTLLKS  
KYIKESVENDIEFAQGISYIEKVLAKYKDDLESIKKVIKEKEPSPPTPPSPA  
KTDQKRESFLPTNIEYLNINLVNKIDYILNKAINDCNVEKDEAHRKIKILS  
DLKAIIDKIDFNTNDFEAIKLLINDDTKMDLKLSTGLVQNFNTIISKLEIK  
FQMLNLSHQCVKQCPENSGCFRHLDERECKLNYKQEGDKCVENPNPCNENN  
GGGDADATCTEEDSGSRKKTICETKPDSPYPLFDGIFCSSNLFGLISFLILMLILY  
SFI"

BASE COUNT 740 a 233 c 245 g 508 t  
ORIGIN

## Alignment Scores:

Pred. No.: 1.23e-126 Length: 1726  
Score: 1931.00 Matches: 373  
Percent Similarity: 96.45% Conservatives: 7  
Best Local Similarity: 94.67% Mismatches: 8  
Query Match: 84.88% Indels: 6  
DB: 3 Gaps: 2

US-10-057-531A-2 (1-431) x PFAMSP843 (1-1726)

Qy 38 ThrAspAspAspLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMet 57  
Db 486 TCCGAAGATAATGCGAATATTATAGATCAAGTAGTAACTGGAGAA----- 530  
Qy 58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77  
Db 531 GCAATATCTGCAATGGAATATATCTCTCAGGATTTGAAATGAATATGATGTTATA 590  
Qy 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97  
Db 591 TATTAAACCTTTAGCTGGAGTATATAGAGCTTAAAAAACAATTTGAAAAACATT 650  
Qy 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117  
Db 651 ATTACATTAATTTAAATTTGACGATATCTTAAATTCACGCTCTTAAGAAACGAAATAT 710  
Qy 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137  
Db 711 TTCTTAGATGATTAGAACTCGATTATATGCAATTTAAACATATATCTCAATGAATAC 770  
Qy 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157  
Db 771 ATTATTGAAGATTTCATTTAAATTTAGATTCAAGAACAAAAACACACATTTTAAAAAGT 830  
Qy 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluLysSer 177

Db 831 TACAAATATATAAAGAATCAGTAGAAAAATCATATTAATTTGCACAGAGGTATAAGT 890  
Qy 178 TyrTyrGluLysValLeuAlaLysTyrIlyAspAspLeuGluSerIleLysLysValIle 197  
Db 891 TATTATGAAAGGTTTGTAGCGAAATATAGGATGATTTAGAAATCAATTTAAAAAGTTATC 950  
Qy 198 LysGluGluLysGluLysPheProSerSerProThrThrProProSerProAlaLys 217  
Db 951 AAAGAAGAAAAGGAG--TTCCCATCATCACCACCAACAACACCTCCGTACCAGCAAAA 1007  
Qy 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237  
Db 1008 ACAGACGAAACAAAAGAGAAAGTAAGTTCTTCATTTTAAACAAACATTGAGACCTTA 1067  
Qy 238 TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn 257  
Db 1068 TACAATAACTTAGTTAATAAATAATGACGATTACTTAATTAACCTTAAGGCAAGATTAAAC 1127  
Qy 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277  
Db 1128 GATTGTAATGTTGAAAGAGATGAACACATCTTAAATAACTAACTTAGTGAATTAAAA 1187  
Qy 278 AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys 297  
Db 1188 GCAATTGATGACAAAATAGATCTTTTAAAAAACACTAACGACCTTCGAAGCAATTTAAAA 1247  
Qy 298 LeuIleAsnAspAspThrLysLysAspMetLeuGluLysLeuLeuSerThrGlyLeuVal 317  
Db 1248 TTGATAATGATGATACGAAAAAAGATATGCTTGGCAANTTACTTAGTACAGGATTAGT 1307  
Qy 318 GlnAsnPheProAsnThrIleIleSerLysLeuIleGluLysPheGlnAspMetLeu 337  
Db 1308 CAAAAATTTTCTCTAATAACAATAATCAAAATTAATTAAGGCAAAATTCGAAGATATGTTA 1367  
Qy 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357  
Db 1368 AACATTTTCAACAACCAATGCGTAAAAAACAATGTCAGAAAAATTTCTGGATGTTTCAGA 1427  
Qy 358 HisLeuAspGluAtyGluGluCysLysCysLeuLeuAsnTyrLysGlnGluLysAspLys 377  
Db 1428 CATTTAGATGAAGAGAAAGATGTAATGTTTAAATTTACAAACAAGAGGTGATAAA 1487  
Qy 378 CysValGluAsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAla 397  
Db 1488 TGTCTTGAATAATCCAAATCCTTACTTGTAAACGAAAAAATATGCTGGATGTCAGATGCC 1547  
Qy 398 ThrCysThrGluGluAspSerGlySerSerArgLysLysIleThrCysGluCysThrLys 417  
Db 1548 ACATGTCACCGAAGAGATTCAGGTAGCAGCAGAAAGAAAAATCACATGTCGAATGACTAAA 1607  
Qy 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431  
Db 1608 CCTGATCTTATCCACATCTTTTCGATGGTATTTTCTGCAGTTCC 1649

## RESULT 17

PFAMSP946

LOCUS

DEFINITION

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

1726 bp DNA linear INV 29-MAY-2002  
P. falciparum DNA for the precursor to the major merozoite surface proteins, C-terminal.

DI3354.1 GI:391818  
EGF-like domains; major merozoite surface protein precursor.  
Plasmodium falciparum (isolate 946) merozoite, DNA.

Plasmodium falciparum  
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
1 (bases 1 to 1726)  
Jongwutiwes,S., Tanabe,K. and Kanbara,H.

Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSPI) of Plasmodium falciparum from field isolates  
Mol. Biochem. Parasitol. 59 (1), 95-100 (1993)  
93293445







TITLE Direct Submission  
JOURNAL Submitted (07-FEB-1995) Ya-Ping Shi, DPD/CDC 4770, Molecular Vaccine Section, Bldg 22, Rm. 8F12, Buford Highway, Chamblee, GA, 30341 USA

FEATURES  
source Location/Qualifiers  
1. .1119  
/organism="Plasmodium falciparum"  
/isolate="infected human blood sample from Papua New Guinea"  
/db\_xref="taxon:5833"  
/clone="B17D"  
1. .1119  
/gene="MSP-1"  
/c1.>.1119  
/gene="MSP-1"  
/codon\_start=1  
/product="merozoite surface protein 1, 42 kDa C-terminal region"  
/protein\_id="AA62224.1"  
/db\_xref="GI:664868"  
LNLNLSRLKRRVFLDVLDFSLDMOFKHSSNEVLIIDFSKLLNSQKNTLLKSYK  
YKESVENDIKFAQEGISYEVKVLAKYKDDLESIKKVIKEEKEPSPPTPPSPAK  
TDEQKESKFLPFLNIEFLYNNLVNKIDDDYLINLAKINDCNVEKNEAHVITILSD  
LKAIDDKIDLFKNTDFEAIKKLINDTKDKMLGKLLSTGLVONFPNTIISKLIEGK  
QDMLNISQHCYKQCPENSGCFRHLDERECKLLNYKQEGDKCVENPDPTCENNG  
GCDADATQTEEDSGSRKKTCECTKPDSPYLFDFGIFCS"

BASE COUNT 487 a 146 c 162 g 324 t

ORIGIN

Alignment Scores:  
Pred. No.: 1.27e-126 Length: 1119  
Score: 1928.00 Matches: 368  
Percent Similarity: 99.20% Conservative: 2  
Best Local Similarity: 98.66% Mismatches: 3  
Query Match: 84.75% Indels: 0  
DB: 3 Gaps: 0

US-10-057-531A-2 (1-431) x PFU20733 (1-1119)

QY 58 AlaileSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77  
|||||  
DB 1 GCAATATCTGTCACATGATGATAATATCTCTCAGGATTTGAAATGAATGATGTTATA 60  
|||||  
QY 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97  
|||||  
DB 61 TATTTAAACCTTTAGCTGAGTATATAGAGCTTAAACAAACAAATTTGAAAAACATT 120  
|||||  
QY 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117  
|||||  
DB 121 ATTACATTTAAATTTAAATTTGAACGATATCTTTAAATTCAGCTTTAAGAAACGAAAAATAT 180  
|||||  
QY 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137  
|||||  
DB 181 TTCCTTAGATGATTAGAACTGATTTAAATGCAATTTAAACATATATCCCAAAATGAATAC 240  
|||||  
QY 138 IleileGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157  
|||||  
DB 241 ATTATTGAAGATTCATTTAAATTTATTTGAATTCAGACAAAAACACACTTTTAAAGAT 300  
|||||  
QY 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177  
|||||  
DB 301 TACAAATATATAAAGAATCAGTAGAAAATGATATTAAATTTTCACAGGAAGGTATAAGT 360  
|||||  
QY 178 TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysValIle 197  
|||||  
DB 361 TATTATGAAGAGGTTTTAGCGAAATATAAGGATGATTGAATCAATTAATAAAGGTATTC 420  
|||||  
QY 198 LysGluGluLysGluLysPheProSerSerProProThrThrProProSerProAlaLys 217  
|||||  
DB 421 AAAGAAGAAAAGGAGAGGTTCCCATCATCATCACCAACAAACACCTCCGTCACCAAGAAA 480  
|||||  
QY 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237  
|||||

Db 481 ACAGACGAACAAAAGAAAGAACTAAGTTCCTCCATTTTAAACAAACATTGAGACCTTA 540  
|||||  
QY 238 TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn 257  
|||||  
Db 541 TACAATAAAGTAACTAATAAATGAGATTAATTAATTAACATAAGGCAAGATTAAC 600  
|||||  
QY 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277  
|||||  
Db 601 GATTGTAAATGTTGAAAAAATGAAGCACATGTTAAATAACTAACTTAGTATTAAAA 660  
|||||  
QY 278 AlaileAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys 297  
|||||  
Db 661 GCAATTGATGACAAAAATAGATCTTTTAAAAACACTAACGACTTCGAAGCAATTAATA 720  
|||||  
QY 298 LeuileAsnAspAspThrLysLysAspMetLeuGlyLysLeuSerThrGlyLeuVal 317  
|||||  
Db 721 TTGATAAATGATGATACGAAAAAAGATATGCTTTGGCAAAATTACTTATGACAGGATTAGT 780  
|||||  
QY 318 GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeu 337  
|||||  
Db 781 CAAAAATTTCCCTAATAATAATATCAAAATTAATTTGAAGAAAAATTTCCAAGATATGTTA 840  
|||||  
QY 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357  
|||||  
Db 841 ACATTTTCAACACCAATCGGTAAAAAACAATGTCCAGAAAAATTTCTGGATGTTTCAGA 900  
|||||  
QY 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377  
|||||  
Db 901 CATTTAGATGAAAGAGAAGAAATGTAATGTTTAAATTTACAAACAAGGAAGGTGATAAA 960  
|||||  
QY 378 CysValGluAsnProAsnProThrCysAsnGluAsnGlyGlyCysAspAlaAspAla 397  
|||||  
Db 961 TGTGTTGAAATTCAGATCTCTACTTGTACGAAAAAATTAATGTTGATGTCAGATGCC 1020  
|||||  
QY 398 ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417  
|||||  
Db 1021 ACATGTACCGAAGAAGATTTCAGGTAGCAGCAAGAAAAATCATCATGTGAATGACTAAA 1080  
|||||  
QY 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSer 430  
|||||  
Db 1081 CCTGATTCCTTATCCACTTTTCGATGATGTTATTTCTGCAGT 1119  
|||||

RESULT 20  
PFU20729 1119 bp DNA linear INV 14-FEB-1995  
LOCUS Plasmodium falciparum clone K47D merozoite surface protein 1  
DEFINITION (MSP-1) gene, 42 kDa C-terminal region, partial cds.  
ACCESSION U20729  
VERSION U20729.1 GI:664859  
KEYWORDS malarial parasite.  
SOURCE Plasmodium falciparum  
ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
REFERENCE Shi,Y.-P., Alpers,M.P., Povoa,M.M., Nahlen,B.L., Oloo,A.G. and Lal,A.A.  
AUTHORS Sequence of the C-terminal region of merozoite surface protein-1 (MSP-1) in field-derived Plasmodium falciparum  
TITLE Unpublished  
JOURNAL 2 (bases 1 to 1119)  
REFERENCE Shi,Y.-P.  
AUTHORS Direct Submission  
TITLE Submitted (07-FEB-1995) Ya-Ping Shi, DPD/CDC 4770, Molecular Vaccine Section, Bldg 22, Rm. 8F12, Buford Highway, Chamblee, GA, 30341 USA  
JOURNAL Location/Qualifiers  
1. .1119  
/organism="Plasmodium falciparum"  
/isolate="infected human blood sample from Papua New Guinea"  
/db\_xref="taxon:5833"  
/clone="K47D"

```
gene      1. .1119
CDS       <1. .>1119
          /gene="MSP-1"
          /gene="MSP-1"
          /codon_start=1
          /product="merozoite surface protein 1, 42 kDa C-terminal
          region"
          /protein_id="AAAG2220.1"
          /db_xref="GI:664860"
          /translation="AISVTMDNLSGFENEYDVYILKPLAGVYRSLKQIEKNLITFN
          LNLNDLNSRLKRRKVELDVEDLMQPKHISNEYIIESFKLLNSEQNTLLKSVK
          YIKESVENDIKFAQEGISYEKVLAKYKDDLESIKKVIKEKEKFPSPPTPPSPAK
          TDQKREKSELPETNLETLYNNLNKIDYLLNLKAKINDCNVEDKDEAHVTKLSD
          LKAIIDDKLKHNDPEAIKLLINDTKKMLGKLSTGLVQNFNPNTIISKIEGKF
          QDMLNIHQVCVKKQCPQNSGCFRHLDEREECKLLNYKQEGDKCVENPNPTCNENG
          GCDADAKCTEEDSGNKKITCECTKPSDYPLFDGIFS"
BASE COUNT 487 a 146 c 161 g 325 t
ORIGIN

Alignment Scores:
Pred. No.:      2,41e-126      Length:      1119
Score:          1924.00      Matches:      367
Percent Similarity: 98.93%      Conservative: 2
Best Local Similarity: 98.39%      Mismatches: 4
Query Match:      84.57%      Indels:      0
DB:              3      Gaps:      0

US-10-057-531A-2 (1-431) x PFU20729 (1-1119)

Qy 58 AlalieserValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValile 77
    |||||
Db 1 GCATATCTGTCAATGGATTAATCTCTCAGGATTTGAAATGAATATGATGTTATA 60

Qy 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97
    |||||
Db 61 TATTAAACCTTTAGCTGGAGTATATAGAGCTTAAAAACAAATGAAAAACATT 120

Qy 98 PheThrPheAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117
    |||||
Db 121 TTTACATTTAATTTAAATTTGAACGATATCTTAAATTCACGCTCTTAAGAAACGAAATAT 180

Qy 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137
    |||||
Db 181 TTCCTAGATGTATAGATCTGATTTAATGCAATTTAAACATATATCTCTCAATGAATAC 240

Qy 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157
    |||||
Db 241 ATTATTGAAGATTTCATTTAAATTTATGAATTCAGAACAAAAACACACTTTTAAAAAGT 300

Qy 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177
    |||||
Db 301 TACAAATATATAAAGAAATCAGTAGAAATGATATTAATTTGCACAGGAAGGTATAAGT 360

Qy 178 TyrTyrGluLysValLeuAlaLysTyrLysAspLeuGluSerIleLysLysValile 197
    |||||
Db 361 TATTATGAAAGGTTTATAGCAATATATAAGATGATTTAGATCAATCAATTAATAAAGTTATC 420

Qy 198 LysGluGluLysGluLysPheProSerSerProProThrThrProProSerProAlaLys 217
    |||||
Db 421 AAAGAAGAAAAAGGAGAGTTCCTCATCATCACCACCAACACACCTCCGTCACCAAGCAAAA 480

Qy 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237
    |||||
Db 481 ACACAGCAACAAAAAGAGAAAGTAAGTTCCCTTCATTTTAAACAAACATTGAGACCTTA 540

Qy 238 TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn 257
    |||||
Db 541 TACATAACTTACTTATAAATTTGACGATTTACTTAATTAATTAAGGCAACATTAAC 600

Qy 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277
    |||||
Db 601 GATTGTAATGTTGAAAAAGATGAAGACATGTTAAAAATAACATAAATCTAGTGATTAAAA 660
```

```
Qy 278 AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys 297
    |||||
Db 661 GCAATTGATGACAAATAGATCTTTTTAAAAACCATACGACTTCGAAGCAATTAATAAAA 720

Qy 298 LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal 317
    |||||
Db 721 TTGATAAATGATGATACGAAAGAGATATGCTTGCAAAATTTACTTAGTACAGGATTAGTT 780

Qy 318 GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGluAspMetLeu 337
    |||||
Db 781 CAAAATTTCTTAATACAAATATCAAAATTAATTAAGGAAAAATCCAAGATATGTTA 840

Qy 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357
    |||||
Db 841 AACATTTCAACAACCAATGCTAAAAAACAATGTCCACAAAATCTCGATGTTTCAGA 900

Qy 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377
    |||||
Db 901 CATTTAGATGAAGAGAGAAATGTAATGTTTAAATTTACAACAAGAGGTGATAAA 960

Qy 378 CysValGluAsnProAsnProThrCysAsnGluAsnGlyGlyCysAspAlaAspAla 397
    |||||
Db 961 TGTGTTGAAATCCAAATCCTACTTGTAAACGAAAAATTAATGTTGGATGTCAGATGCC 1020

Qy 398 ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417
    |||||
Db 1021 AAATGTCACCGAAGAGATTTCAGTAGCAACGGAAGAAAAATCACATGTGAATGTACTAAA 1080

Qy 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSer 430
    |||||
Db 1081 CCTGATCTTATCACCCTTTTCGATGGTATTTTCGCGAGT 1119

RESULT 21
PFU20653
LOCUS      PFU20653              1116 bp      DNA      linear      INV 14-FEB-1995
DEFINITION Plasmodium falciparum clone P24B merozoite surface protein 1
            (MSP-1) gene, 42 kDa C-terminal region, partial cds.
ACCESSION U20653
VERSION    020653.1
KEYWORDS   GI:664843
SOURCE     malarial parasite.
ORGANISM   Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE  1 (bases 1 to 1116)
            Shi, Y.-P., Alpers, M.P., Povoa, M.M., Nahlen, B.L., Oloo, A.G. and
            Lal, A.A.
            Sequence of the C-terminal region of merozoite surface protein-1
            (MSP-1) in field-derived Plasmodium falciparum
            Unpublished
            2 (bases 1 to 1116)
            Shi, Y.-P.
            Direct Submission
            Submitted (06-FEB-1995) Ya-Ping Shi, Molecular Vaccine Section,
            Bldg 22, Rm. 8F12, DPD/CDC 4770, Buford Highway, Chamblee, GA 30341
FEATURES
            source
            1. .1116
            /organism="Plasmodium falciparum"
            /isolate="infected human blood sample from Papua New
            Guinea"
            /db_xref="taxon:5833"
            /clone="P24B"
            1. .1116
            /gene="MSP-1"
            <1. .>1116
            /gene="MSP-1"
            /codon_start=1
            /product="merozoite surface protein 1, 42 kDa C-terminal
            region"
            /protein_id="AAAG2220.1"
            /db_xref="GI:664844"
            /translation="AISVTMDNLSGFENEYDVYILKPLAGVYRSLKQIEKNLITFN
            LNLNDLNSRLKRRKVELDVEDLMQPKHISNEYIIESFKLLNSEQNTLLKSVK
            YIKESVENDIKFAQEGISYEKVLAKYKDDLESIKKVIKEKEKFPSPPTPPSPAK
```

[illegible]

PDVTPSLSV

CDS



LKAIDDKIDLKFNHNDFAIKKLIINDTKKMLGKLLSTGLVQNPNTIISKLEIGKF  
 ODLNLSQHYVKKOCPENSGCFRHLDERECKCLLNKQEGDKCVENPNTCENNG  
 GCDADAKTEEDSGSKKITCECTKPDSPFLDFGIFS"

BASE COUNT 487 a 145 c 162 g 325 t  
 ORIGIN

## Alignment Scores:

Pred. No.: 5,41e-126 Length: 1119  
 Score: 1919.00 Matches: 368  
 Percent Similarity: 98.66% Conservative: 0  
 Best Local Similarity: 98.66% Mismatches: 5  
 Query Match: 84.35% Indels: 0  
 DB: 3 Gaps: 0

US-10-057-531A-2 (1-431) x AF040567 (1-1119)

QY 58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77  
 Db 1 GCAAATATCTGTCAAAATGGATAATATCTCTCAGGATTTGAAATGAATGATGATTATA 60  
 QY 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97  
 Db 61 TATTTAAACCTTTACCTGGATATATAGAGCTTAAAAAACAATTTGAAAAACATT 120  
 QY 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117  
 Db 121 TTTACATTTTAAATTTGAACGATATCTTTAAATTCAGCTTAAAGAAACGAAAAATAT 180  
 QY 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137  
 Db 181 TTTCTTAGATGATTAAGATCTGATTAATGCAATTTAAACATATATCTTCAATGAATAC 240  
 QY 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157  
 Db 241 ATTATTGAAGATTCATTTAAATATTGAATTCAGACAAAAACACACTTTTAAAGT 300  
 QY 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177  
 Db 301 TACAAATATATAAAGAAATCAGTAGAAAATGATATTAAATTTGCACAGGAAGGTATAAGT 360  
 QY 178 TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValIle 197  
 Db 361 TATTATGAAGAGTTTACGGAATATAGGATGATTAGATCAATTTAAAAAGTTATC 420  
 QY 198 LysGluGluLysLysPheProSerSerProProThrThrProProSerProAlaLys 217  
 Db 421 AAAGAAGAAAGGAGAAGTTCCCATCATCACCAACCAACACCTCCGTCACCCAGCAAAA 480  
 QY 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237  
 Db 481 ACAGACGACAAAGAAAGGAAAGTAAAGTTCCCTTCCATTTTAAACAAACATTTGAGACCTTA 540  
 QY 238 TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn 257  
 Db 541 TACAATAACTTAGTTAATAAATGACGATTAATTAACCTTAAGGCAACAAATTAAC 600  
 QY 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277  
 Db 601 GATTGTAGTCTTGAAAGATGAGACACATGTTAAAAATACTAACTTAGTGATTTAAAA 660  
 QY 278 AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys 297  
 Db 661 GCAATTTGATGACAAAAATAGATCTTTTAAAAACCATTAAGCACTTCAAGCAATTAATAAAA 720  
 QY 298 LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal 317  
 Db 721 TTGATTAATGATGATACGAAAAAAGATATGCTTTGGCAAAATTACTTAGTACAGGATTAGTT 780  
 QY 318 GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeu 337  
 Db 781 CAAAAATTTCTTAATACATAATATCAAAATTAATTTGAAGGAAAAATTTCCAGATATGTTA 840  
 QY 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357

Db 841 AACATTTCAACACCAATACGTAAAAAACAATGTCCAGAAAATCTCGATGTTTCAGA 900  
 QY 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377  
 Db 901 CATTTAGATGAAAGAGAAGAAATGTAATGTTTATTAATTAACAAACAAGAGTGATAAA 960  
 QY 378 CysValGluAsnProAsnProThrCysAsnGluAsnGlyGlyCysAspAlaAspAla 397  
 Db 961 TGTGTTGAAAATCCAATCCTACTTGTACGAAAAAATAATGTTGATGATGATGATGCC 1020  
 QY 398 ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417  
 Db 1021 AAATGTACCGAAGAAGATTACGTAGCAGCGGAAAGAAATACATCTGTAATGATCTACTAAA 1080  
 QY 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSer 430  
 Db 1081 CTTGATTTCTTATCCACTTTTCGATGATGATTTCTTCGAGT 1119  
 RESULT 24  
 PFU20726 1119 bp DNA linear INV 14-FEB-1995  
 LOCUS Plasmodium falciparum clone K46A merozoite surface protein 1  
 DEFINITION (MSP-1) gene, 42 kDa C-terminal region, partial cds.  
 ACCESSION U20726  
 VERSION U20726.1 GI:664853  
 KEYWORDS malarial parasite.  
 SOURCE Plasmodium falciparum  
 ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 REFERENCE 1 (bases 1 to 1119)  
 AUTHORS Shi,Y.-P., Alpers,M.P., Povoa,M.M., Nahlen,B.L., Oloo,A.G. and Lal,A.A.  
 TITLE Sequence of the C-terminal region of merozoite surface protein-1 (MSP-1) in field-derived Plasmodium falciparum  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 1119)  
 AUTHORS Shi,Y.-P.  
 TITLE Direct Submission  
 JOURNAL Submitted (07-FEB-1995) Va-Ping Shi, DPD/CDC 4770, Molecular Vaccine Section, Bldg 22, Rm. 8F12, Buford Highway, Chamblee, GA, 30341 USA  
 FEATURES  
 source Location/Qualifiers  
 1..1119 /organism="Plasmodium falciparum"  
 /isolate="infected human blood sample from Papua New Guinea"  
 /db\_xref="taxon:5833"  
 /clone="K46A"  
 1..1119 /gene="MSP-1"  
 <1..>1119 /gene="MSP-1"  
 /codon\_start=1 /product="merozoite surface protein 1, 42 kDa C-terminal region"  
 /protein\_id="AAA62217.1"  
 /db\_xref="GI:664854"  
 /translation="ALSVTMDNLTLSGFENEYDVIYLLKPLAGVYRSLLKQIEKNIFTFN LNLNLLNSRLKKRYFLDVLSDLMQFKHISSENYIIEDSFKLLNSEKNTLLKSYK YIKESVENDIKFAQEGISYIERLYAKYKDDLSIRKVIKEEKFPSPPTTPPSPAK TDEQKESKFLPFLNIETLYNNLVNKIDHYLNKAKINDCNVERDEARVKTLSL LKAIIDKIDLFKNHNDFAIKKLIINDTKKMLGKLLSTGLVQNPNTIISKLEIGKF ODLNLSQHYVKKOCPENSGCFRHLDERECKCLLNKQEGDKCVENPNTCENNG GCDADAKTEEDSGSKKITCECTKPDSPFLDFGIFS"  
 BASE COUNT 486 a 146 c 162 g 325 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 5,41e-126 Length: 1119  
 Score: 1919.00 Matches: 366  
 Percent Similarity: 98.66% Conservative: 2  
 Best Local Similarity: 98.12% Mismatches: 5

```
Query Match: 84.35% Indels: 0
DB: 3 Gaps: 0
US-10-057-531A-2 (1-431) x PFU20726 (1-1119)

Qy 58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77
Db 1 GCAATATCTGTCACAAATGGATTAATCTCTCAGGATTGAAATGAATGATGTTATA 60

Qy 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysGlnIleGluLysAsnIle 97
Db 61 TATTTAAACCTTTAGCTGGAGTATATAGAGCTTAAAAAACAATTTGAATAAACAATT 120

Qy 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117
Db 121 TTTACATTTAATTTAAATTTCAACGATATCTTAAATTCACGCTCTTAAGAAACGAAATAT 180

Qy 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137
Db 181 TTCTTAGATGTTATAGAAATCTGATTTAATGCAATTTAAACATATATCTCAATGAATAC 240

Qy 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLysSer 157
Db 241 ATTATTGAAGATTCATTTAAATTTAATTTAGATTCAGAACAAAAACACACTTTTAAAAAGT 300

Qy 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177
Db 301 TACAATATATAAAGAAATCAGTAAATATGATTTAAATTTGCACAGGAAGGTATAAGT 360

Qy 178 TyrTyrGluLysValLeuAlaLysTyrLysAspLeuGluSerIleLysLysValIle 197
Db 361 TATTATGAAAGGTTTACGCAATATATAGGATGATTTAGATCAATTTAAAAAGTTATC 420

Qy 198 LysGluGluLysGluLysPheProSerProProThrThrProProSerProAlaLys 217
Db 421 AAAGAAGAAAGGAGAGAGTCCCATCATCATCACCAACACACCTCCGTCACAGCAAAA 480

Qy 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237
Db 481 ACACGACGACAAAAGAGAGAAAGTAAAGTTCCTCATTTTAAACAAACATTTGAGACCTTA 540

Qy 238 TyrAsnAsnLeuValAsnLysIleAspTyrLeuIleAsnLeuLysAlaLysIleAsn 257
Db 541 TACAATACTTAGTTAATAAATTAAGCATTTACCATTTAATTAATTAAGGCAAGGATTAAC 600

Qy 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277
Db 601 CATTTGTAATGTTGAAAGAGATGAGCACATGTTAAATAACTAACTTAGCTGATTTAAA 660

Qy 278 AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys 297
Db 661 GCAATTGATGACAAAATAGATCTTTTAAAAAACCATACGACTTCGAAGCAATTTAAAAA 720

Qy 298 LeuIleAsnAspAspThrLysLysAspMetLeuGluLysLeuLeuSerThrGlyLeuVal 317
Db 721 TTGATAATGATGATACGAAAAAGATATGTTGGCAATTTACTTAGTACAGGATTAGTT 780

Qy 318 GlnAsnPheProAsnThrIleIleSerLysLeuIleGluLysPheGlnAspMetLeu 337
Db 781 CAAATTTTCTTAAATACATAATATCAAAATTAATTAAGGAAATTTCCAAGATATGTTA 840

Qy 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357
Db 841 AACATTTTCAACAACCAATGCGTAAAAAACAATGTCCAGAAAAATTTCTGGATGTTTCAGA 900

Qy 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluLysAspLys 377
Db 901 CATTTAGTAAAGAGAAAGATGTAATGTTTATTAATTAACAACAGAAAGGTTGATAA 960

Qy 378 CysValGluAsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAla 397
Db 961 TGTGTTGAAATCCAAATCCTTACTTGTAAACGAAATTAATGTTGATGTCAGATGCC 1020
```

```
Qy 398 ThrCysThrGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417
Db 1021 AAATGTACCGAAGAAGATTACAGTAGCAACGAAAGAAAGTACATGTGAATGACTATAA 1080

Qy 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSer 430
Db 1081 CCTGATTCTTATCCACTTTTCGATGGTATTTTTCGAGT 1119

RESULT 25
PFU20654 1116 bp DNA linear INV 14-FEB-1995
LOCUS Plasmodium falciparum clone P24C merozoite surface protein 1
DEFINITION (MSP-1) gene, 42 kDa C-terminal region, partial cds.
ACCESSION U20654
VERSION U20654.1 GI:664845
KEYWORDS malarial parasite.
SOURCE Plasmodium falciparum
ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 1116)
AUTHORS Shi,Y.-P., Alpers,M.P., Pova,M.M., Nahlen,B.L., Oloo,A.G. and Lal,A.A.
TITLE Sequence of the C-terminal region of merozoite surface protein-1 (MSP-1) in field-derived Plasmodium falciparum
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1116)
AUTHORS Shi,Y.-P.
TITLE Direct Submission
JOURNAL Submitted (06-FEB-1995) Ya-Ping Shi, Molecular Vaccine Section, Bldg 22, Rm. 8F12, DPD/CDC 4770, Buford Highway, Chamblee, GA 30341
FEATURES
    source
        location/Qualifiers
            1..1116
                /organism="Plasmodium falciparum"
                /isolate="infected human blood sample from Papua New Guinea"
                /db_xref="taxon:5833"
                /clone="P24C"
                /gene="MSP-1"
                /gene="MSP-1"
                /codon_start=1
                /product="merozoite surface protein 1, 42 kDa C-terminal region"
                /protein_id="AAAG62214.1"
                /db_xref="GI:664846"
                /translation="ATSVTMDNLLSGFENEVDVLYLKPLAGVYRSKLKQLEKNLIITEN
                LNLNDILNSRLKRRKFLDLESDLMQFKHISSEYIIESDKKLEKEFPSPPTPPSPAKT
                YIKESVENDIKFAQEGISYYEKVLAKYDDLESIKKVIKEKEFPSPPTPPSPAKT
                DEQKESKFLPTNITETLYNNLVNKIDDDYLINLKAINDCNVEKDEAHYKIPKLSDL
                KATDDKIDLEKNTNDFAIKKLLINDDTKDLMLGLKLTSLGLVQNPNTIISKLEKQFO
                DMLNISQHCVKKQCPENSGCFRHLDERECCLLNLYKQEGDKCVENSPSTCENNGG
                CDADATCTEEDSGSRKKITCECTKPPSYPLFDGIFCS"

BASE COUNT 484 a 145 c 162 g 325 t
ORIGIN

Alignment Scores:
Pred. No.: 6.87e-126 Length: 1116
Score: 1917.50 Matches: 368
Percent Similarity: 98.93% Conservative: 1
Best Local Similarity: 98.66% Mismatches: 3
Query Match: 84.29% Indels: 1
DB: 3 Gaps: 1

US-10-057-531A-2 (1-431) x PFU20654 (1-1116)

Qy 58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77
Db 1 GCAATATCTGTCACAAATGGATTAATCTCTCAGGATTGAAATGAATGATGTTATA 60

Qy 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysGlnIleGluLysAsnIle 97
Db 61 TATTTAAACCTTTAGCTGGAGTATATAGAGCTTAAAAAACAATTTGAATAAACAATT 120
```



QY	98	PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr	117
DB	121	ATTACATTTAAATTTAAATTTGAACGATATCTTAAATTCAGCTTAAAGAAACGAAATAT	180
QY	118	PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr	137
DB	181	TTCTTAGATGTTATGAATCTGATTTAATGCAATTTAAACATATATATCCCAATGAATAC	240
QY	138	IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLysSer	157
DB	241	ATTATTGAAGATTCATTTAAATTTGAACGATATCTTAAATTCAGCTTAAAGAAACGAAATAT	300
QY	158	TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer	177
DB	301	TACAAATATATAAAGAAATCAGTAGAAATGATATTAATTTGACAGCAAGGTATAAGT	360
QY	178	TyrTyrGluValLeuAlaLysTyrLysAspAspLeuGluSerIleLysValIle	197
DB	361	TATTATGAAGAGTTTACGGAATATAAGGATGATTAGATCAATTTAAAGAAAGTTATC	420
QY	198	LysGluGluLysGluLysPheProSerSerProThrThrProProSerProAlaLys	217
DB	421	AAAGAGAAAGAG---TTCCCATCATCACCACCAACACACCTCCGTCACCAACGAA	477
QY	218	ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu	237
DB	478	ACAGACGAAACAAAGAAAGTAAGTTCCTTCCTCCATTTTAAACAAACATTCAGACCTTA	537
QY	238	TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn	257
DB	538	TACAATACTTAGTTAAATAAATTCAGGATTAATTAATTAAGTAAAGCAAGATTAAC	597
QY	258	AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys	277
DB	598	GATTGTAATGTTGAAAGATGACGACATGTTAAATAACTTAACCTTACTGATTTAA	557
QY	278	AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys	297
DB	658	GCAATTTGATGACAAATAGATCTTTTAAACACACTAACGACTTCCGAAGCAATTA	717
QY	298	LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal	317
DB	718	TTGATAAATGATGATACGAAAGAAAGATATGCTTGGCAATTAATCTAGTACAGGATG	777
QY	318	GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeu	337
DB	778	CAAAATTTTCTTAATACATAATATATCAAAATTAATGAAGAAATTTCCACAGATGTTA	837
QY	338	AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg	357
DB	838	AACATTTTCAACACCAATGCGTAAACAAACAAATGTCCAGAAATTTCTGGATCTTCA	897
QY	358	HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys	377
DB	898	CATTTAGATCAAGAGAAAGAAATGTAATGTTTATTAAATTAACAAAGAAAGTGATAA	957
QY	378	CysValGluAsnProAsnProThrCysAsnGluAsnGlnGlyCysAspAlaAspAla	397
DB	958	TCTGTTGAAATCCCAAGTCTTACTTGTAAACGAAATAATGTTGGATGTCATGCAGTGC	1017
QY	398	ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys	417
DB	1018	ACATGTACCGAAGAAAGATTCAGGTAGCAGCAAGAAAGAAATCACATGTGAATGTACTAA	1077
QY	418	ProAspSerTyrProLeuPheAspGlyIlePheCysSer	430
DB	1078	CCTGATCTTATCCACTTTTTCGATGTTATTTCTGCAGT	1116
RESULT	26		
PFU20655			
LOCUS			
DEFINITION			

(MSP-1) gene, 42 kDa C-terminal region, partial cds.	
ACCESSION	U20655
VERSION	U20655.1
KEYWORDS	GI:664847
SOURCE	malerial parasite.
ORGANISM	Plasmodium falciparum
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS	1 (bases 1 to 1116) Shi.Y.-P., Alpers.M.P., Povia.M.M., Nahlen,B.L., dloo,A.G. and Lal.A.A.
TITLE	Sequence of the C-terminal region of merozoite surface protein-1 (MSP-1) in field-derived Plasmodium falciparum
REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 1116) Shi.Y.-P.
JOURNAL	Direct Submission
TITLE	Submitted (06-FEB-1995) Ya-Ping Shi, Molecular Vaccine Section, Bldg 22, Rm. 8F12, DPD/CDC 4770, Buford Highway, Chamblee, GA 30341
FEATURES	Location/Qualifiers
source	1..1116 /organism="plasmodium falciparum" /isolate="infected human blood sample from Papua New Guinea" /db_xref="taxon:5833" /clone="P27C" 1..1116 /gene="MSP-1" <1..>1116 /gene="MSP-1" /codon_start=1 /product="merozoite surface protein 1, 42 kDa C-terminal region" /protein_id="AAA62215.1" /db_xref="GI:664848"
gene	/translation="AISTVMDNLSGFENEYDVIYKPLAGVYRSUKKQIEKNITFFN LNLNDLSRLKKRYFDLVDLESLMQFKHSSNEYIIEESFKLNSKQNTLLKSYK YIKESVENDIKPAQEGISYKELVLAQYKDDLESIKKVEKEEPSPPTPPSPAK DEQKESKFLPFLNIETLYNNLVNKIDYILNLKAKINDCNVENEKHAHVITKLSDL KAIIDKIDLFKNTDFEAIKKLINDTKMLKGLSTGLVQNFNTITISKLEGFQ DMLNLSQHCVKKQENSGCFRHLDEREECKLLNYKQEGDKCVENPNPTCNEGG CDADATCTEEDSGSSRKKTCTCEKTPDSYPLFDGIFS"
CDS	
BASE COUNT	486 a 146 c 160 g 324 t
ORIGIN	
Alignment Scores:	
Pred. No.:	6.87e-126 Length: 1116
Score:	1917.50 Matches: 368
Percent Similarity:	98.93% Conservatives: 1
Best Local Similarity:	98.66% Mismatches: 3
Query Match:	84.29% Indels: 1
DB:	3 Gaps: 1
US-10-057-531A-2 (1-431) x PFU20655 (1-1116)	
QY	58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77
Db	1 GCAATATCTGTCCAAATGGATATATCTCTCAGGATTTGAAATGAATATGATGTTATA 60
QY	78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97
Db	61 TATTTAAACCTTTTACTGGAGTATATAGAAGCTTAAAAACAAATTTGAAAAACATT 120
QY	98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117
Db	121 ATTACATTTTAAATTTAAATTTGAACGATATCTTAAATTCAGCTTAAAGAAACGAAATAT 180
QY	118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137
Db	181 TTCTTAGATGTTATAGAACTGATTTAATGCAATTTAAACATATATCTCAATGAATAC 240
QY	138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLysSer 157
Db	241 ATTATTGAAGATTCATTTAAATTTTCAATTCAGAAACAAACAAACACTTTTAAAAAGT 300

Qy	158	TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer	177	Unpublished
Db	301	TACAAATATATAAAGAATCATAGTAAATATATAATTTGCACAGGAGGTATAAGT	360	Shi, Y.-P.
Qy	178	TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValle	197	Direct Submission
Db	361	TATTATGAAAGGTTTGTAGCGAAATATAAGGATGATTAGATCAATTTAAAAAGTTATC	420	Submitted (06-FEB-1995) Ya-Ping Shi, Molecular Vaccine Section, Bldg 22, Rm. 8F12, DPD/CDC 4770, Buford Highway, Chamblee, GA 30341
Qy	198	LysGluLysGluLysPheProSerSerProProThrThrProProSerProAlaLys	217	Location/Qualifiers
Db	421	AAAGAGAAAGGAG---TTCCCATCATCACCACCAACACACCTCCGTCCACGACAAA	477	1. .1116
Qy	218	ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu	237	/organism="Plasmodium falciparum"
Db	478	ACAGCAACAAAAAGAGAAAGTAAAGTTCCTCCATTTTAAACAAACATTGACACCTTA	537	/isolate="infected human blood sample from Papua New Guinea"
Qy	238	TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLysAlaLysIleAsn	257	/db_xref="taxon:5833"
Db	538	TACAAATACCTTAGTTAAATAAATTGACGATTACTTAATTAACTTAAAGGCAAGATTAAAC	597	/clone="P27D"
Qy	258	AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys	277	1. .1116
Db	598	GATTGTAATGTTGAAAAAATGACGACATGTTAAATACTAACTTAGTGATTAAAA	657	/gene="MSP-1"
Qy	278	AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys	297	<1. >1116
Db	658	GCAATTGATGACAAATAGATCTTTTAAAAAACAACACTAACGACTTCGAAGCAATTA	717	/gene="MSP-1"
Qy	298	LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuSerThrGlyLeuVal	317	/codon_start=1
Db	718	TTGATAATGATGATACGAAAAAAGATGCTTGGCAATTTACTTAGTACAGGATTAGTT	777	/product="merozoite surface protein 1, 42 kDa C-terminal region"
Qy	318	GlnAsnPheProAsnThrIleIleSerLysLeuIleGluLysPheGlnAspMetLeu	337	/protein_id="AAA62216.1"
Db	778	CAAAATTTCTCTAATACATAATATCAAAATTAATTGAAGCAAAATTCAGATATGTTA	837	/db_xref="GI:664850"
Qy	338	AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg	357	/translation="ATSVTMDNITLSCFENEYDVIYLPKLAGVYRSLLKQIEKNITFN LNNDILNRLKRRKIFLDVSDLMQFQKHISSEYTIEDSFRLNSEQNRLKSYK YIKRESVNDIKFAQEGISYIEKVLAKDDLESIKKVIKEKEFPSPPTPPSPAKT DEQKESKFLPFLNTIETIYNLNKIDDDYILNLKAKINDCNVKEAHVKITKLSDL KATDKRIDLFKNTNDFEAIKKLINDDTKMDLKLKLLSTGLVQNPNTIISKLEGFQ DMLNISOHCVCKKCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCENNGG CDADATCTEEDSGSSRRKICTCECTKPSYFLFDGIFS"
Db	838	AACATTTCAACAACCAATGCGTAAAAAACAATGTCAGAAAAATTCGGATGTTTCAGA	897	BASE COUNT 486 a 146 c 160 g 324 t
Qy	358	HisLeuAspGluAtgGluGluCysLysCysLeuLeuAsnTyrLysGlnGlyAspLys	377	ORIGIN
Db	898	CATTAGATGAAGAGAGAAATGTAATGTTTATTAATTAACAACGAAGAGGTGATAA	957	Alignment Scores:
Qy	378	CysValGluAsnProAsnProThrCysAsnGluAsnGlyGlyCysAspAlaAspAla	397	Pred. No.: 6.87e-126 Length: 1116
Db	958	TGCTGTAATAATCCAAATCCTACTTGTAAACGAAATAATGTTGATGTCAGATGCC	1017	Score: 1917.50 Matches: 368
Qy	398	ThrCysThrGluGluAspSerGlySerSerArgLysLysIleThrCysGluCysThrLys	417	Percent Similarity: 98.93% Conservativeness: 1
Db	1018	ACATGTACCGAAGAGATTTCAGGTAGCAGCAGAAAGAAAAATCACATGTGAATGTACTAA	1077	Best Local Similarity: 98.66% Mismatches: 3
Qy	418	ProAspSerTyrProLeuPheAspGlyIlePheCysSer	430	Query Match: 84.29% Indels: 1
Db	1078	CCTGATCTTATCCATTTTCGATGGTATTTTCGCAGT	1116	DB: 3
RESULT 27				US-10-057-531A-2 (1-431) x PFU20656 (1-1116)
PFU20656				
LOCUS				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				



Qy 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177  
 Db 1075 TACAATATATAAAGAAATCATAGTAAATGATATTAATTTGCACAGGAAGGTATAAGT 1134  
 Qy 178 TyrTyrGluLysValLeuAlaLysTyrLysAspLeuGluSerIleLysLysValIle 197  
 Db 1135 TATTATGAAAGGTTTGTAGCGAATATAAGGATGATTAGTAATCAATTAATAAAGTATTATC 1194  
 Qy 198 LysGluGluLysGluLysPheProSerSerProProThrThrProProSerProAlaLys 217  
 Db 1195 AAAGAAGAAAGGAG---TTCCCATCATCACCACCAACACACCTCCGTACACGACAAA 1251  
 Qy 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237  
 Db 1252 ACAGACGAACAAAAGAGAAAGTAAAGTCTCTCCATTTTAAACAAACATTGAGACCTTA 1311  
 Qy 238 TyrAsnAsnLeuValAsnLysIleAspAspThrLeuIleAsnLeuLysAlaLysIleAsn 257  
 Db 1312 TCAATAACTTAGTTAATAAATTGACGATTACTTAATTAACCTTAAGGCAAGATTAAAC 1371  
 Qy 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277  
 Db 1372 GATTGTAATGTTGAAAGATGAAGCACATGTTAAATAAATACTAACTTAGTGATTTAAA 1431  
 Qy 278 AlaIleAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys 297  
 Db 1432 GCAATTGATGACAAAATAGATCTTTTAAACCAATACGACTTCGAAAGCAATTTAAAA 1491  
 Qy 298 LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal 317  
 Db 1492 TTGATAATGATGATACGAAAAGATATGCTTGGCAAAATTACTTAGTACAGGATTAGTT 1551  
 Qy 318 GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeu 337  
 Db 1552 CAAAATTTCTTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 1611  
 Qy 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357  
 Db 1612 AACATTTTCAACACCAATGCGTAAAAAACAATGTCACACAAATTCGGATGTTTCAGA 1671  
 Qy 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377  
 Db 1672 CATTTAGTGAAGAGAGATGTAATGTTTATTAATAATAACAACAAGAGGTGATAA 1731  
 Qy 378 CysValGluAsnProAsnProThrCysAsnGluAsnGlyGlyCysAspAlaAspAla 397  
 Db 1732 TGTGTTGAAATCCAAATCCTACTTGTAAACGAAATAATGTTGGATGTCAGATGCC 1791  
 Qy 398 ThrCysThrGluGluAspSerGlySerSerArgLysLysIleThrCysGluCysThrLys 417  
 Db 1792 AAATGTTACCGAAGAGATTCAGGTAGCAACCGAAGAAATAATCACATGTGATGTACTAAA 1851  
 Qy 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431  
 Db 1852 CCTGATCTTATCCACITTTTCGATGGTATTTCCTGCAGTTCC 1893

RESULT 29  
 PFAMSP815  
 LOCUS P. falciparum DNA 1726 bp DNA linear INV 29-MAY-2002  
 DEFINITION P. falciparum DNA for the precursor to the major merozoite surface proteins, C-terminal.  
 ACCESSION D13346  
 VERSION D13346.1 GI:391792  
 KEYWORDS EGF-like domains; major merozoite surface protein precursor.  
 SOURCE Plasmodium falciparum (isolate 815) merozoite, DNA.  
 ORGANISM Plasmodium falciparum  
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 REFERENCE 1 (bases 1 to 1726)  
 JONGWUTIWES.S., TANABE,K. and KANBARA,H.  
 TITLE Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP1) of Plasmodium falciparum from field isolates  
 JOURNAL Mol. Biochem. Parasitol. 59 (1), 95-100 (1993)

93295445  
 2 (bases 1 to 1726)  
 JONGWUTIWES.S.  
 Direct Submission  
 Submitted (05-OCT-1992) Somchai Jongwutiwes, Institute of Tropical Medicine, Nagasaki Univ., Dept. of Protozoology; 12-4 Sakamoto-machi, Nagasaki 852, Japan (Tel:0958-47-2111(ex.3747), Fax:0958-47-6607)  
 Submitted (05-OCT-1992) to DDBJ by:  
 Somchai Jongwutiwes  
 Department of Protozoology  
 Institute of Tropical Medicine  
 Nagasaki University  
 12-4 Sakamoto-machi  
 Nagasaki 852  
 Japan  
 Phone: 0958-47-2111x3747  
 Fax: 0958-47-6607  
 Location/Qualifiers  
 1..1726  
 /organism="Plasmodium falciparum"  
 /db\_xref="taxon:5833"  
 /chromosome="9"  
 /dev\_stage="merozoite"  
 1..1712  
 /gene="MSP1"  
 <1..1712  
 /gene="MSP1"  
 /codon\_start=3  
 /product="major merozoite surface protein precursor"  
 /protein\_id="BAA02607.1"  
 /db\_xref="GI:391793"  
 /translation="PLKTLSEVSIQTDNYANLEKFRVLSKIDGKLNLDNHLGKKKLS  
 FLSSGLHLITELKVKNNKNTGNSPNNKKNVNEALKSYENFLPEAKVTVVTPPQ  
 PDVTPSLSVRVSGSGSTKEETQITPSGLLTQLOVQVOLQNYDEEDSLVLPFG  
 ESDNDDEYLDQVVTGEAISVTMDNILLSGFENEYDVLYLPLAGVRSLLKQIEKNIIT  
 FNUNLDILNSRLKKRKYFLDLESLLMQFKHSSNEYIEDSFKLNSOKNTLLKS  
 YTKIKESVENDIKFAQEGISYEKVLAKYKDDLESKKVKEKEKPESSPTTPPSPA  
 KTDQKESKFLPLNIETLYNNLVNKIDYILNKAKINDCNVEKPEDEAHVKILS  
 DLKAIIDKIDLFKNTDFAIKKLIINDDTKKMLGKLLSTGLVQNPPTTISLIEGK  
 FQMLNLSHQCVKQCPNSGCFRHLDRECKLLNLYKQEGDKGVENPNPTCNENN  
 GGCDAKACTEEDSGSNGKKTCTCECTKPDSPYPLFDGIFCSSSNFLGTSFLILMLLY  
 SFI"  
 BASE COUNT 741 a 232 c 245 g 508 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1..62e-125 Length: 1726  
 Score: 1915.00 Matches: 370  
 Percent Similarity: 95.94% Conservativeness: 8  
 Best Local Similarity: 93.91% Mismatches: 10  
 Query Match: 84.18% Indels: 6  
 DB: 3 Gaps: 2  
 US-10-057-531A-2 (1-431) x PFAMSP815 (1-1726)  
 Qy 38 ThrAspAspAspLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMet 57  
 Db 486 TCCGAAGATAATGACGAATATTAGTCAAGTAGTACCTGGAGAA----- 530  
 Qy 58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77  
 Db 531 GCAATATCTGTCACAAATGGATAATATCTCTCAGGATTTGAAAATGAATATGATGTTATA 590  
 Qy 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysGlnIleGluLysAsnIle 97  
 Db 591 TATTAAACCTTTAGCTGGAGTATATAGAGCTTAAAAACAACAAATGAAAAACATT 650  
 Qy 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117  
 Db 651 ATTACATTTAATTTAAATTTGACGATATCTTAAATTCACGCTCTTAAGAAACGAAATAT 710  
 Qy 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137

Db 711 TTCTAGATGATTAAGTCTGATTAATGCAATTTAAACATATATCTCTCAATGATAC 770  
Qy 138 ileileGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157  
Db 771 ATTATTGAAGATTCATTAATTAATTAATTCAGACAAAAACACACTTTTAAAAAGT 830  
Qy 158 TyLysTyrlLysGluSerValGluAsnAspLysPheAlaGlnGluGlyIleSer 177  
Db 831 TACAAATATATAAAGAATCAGTAGAAAAATGATTAATTAATTCACAGGAAGGTATAAGT 890  
Qy 178 TyTyTyrCluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValIle 197  
Db 891 TATTATGAAAAGGTTTATAGCGAATATAAGGATGATTAGAATCAATTAATAAAGATTATC 950  
Qy 198 LysGluGluLysGluLysPheProSerSerProProThrThrProProSerProAlaLys 217  
Db 951 AAAGAAGAAAAGGAG---TTCCCATCATCACACACACACTCGTCCACGACAAAA 1007  
Qy 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237  
Db 1008 ACAGACGAACAAAAGGAAGGAAAGTAAGTTCCTTCCATTTTAAACAACATTGAGACCTTA 1067  
Qy 238 TyAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn 257  
Db 1068 TACAATAACTTAGTTAATAAAATTTGACGATTACTTAATTAACCTAAAGCGAAGATTAAAC 1127  
Qy 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277  
Db 1128 GATTGTAAGTTGAAAAGATGAGCACATGTTAAATACTAACTTAGTATTAAAA 1187  
Qy 278 AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys 297  
Db 1188 GCAATTGATGACAAAATAGATCTTTTAAACACTAACGACTTCGAAGCAATTAATAAA 1247  
Qy 298 LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal 317  
Db 1248 TTGATAAATGATGATACGAAAAAAGATATGCTGGCAAAATTAATTAACAGGATTAGTT 1307  
Qy 318 GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeu 337  
Db 1308 CAAATTTTCTTAATCAATAATATCAAAATTAATTAAGGAAAAATTCGAAGATATGTTA 1367  
Qy 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357  
Db 1368 AACATTTCAACACCAATCGTAAAAAAACAATGTCCAGAAAATTCGTGATGTTTCA 1427  
Qy 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377  
Db 1428 CATTTAGATGAAGAGAAGATGTAATGTTTATTAATAATTAACAAACAAGAGGTGATAAA 1487  
Qy 378 CysValGluAsnProAsnProThrCysAsnGluAsnGlyGlyCysAspAlaAspAla 397  
Db 1488 TGTGTTGAAATCCAAATCTACTTGTAAAGAAAATTAATGTTGATGTCAGATGCC 1547  
Qy 398 ThrCysThrGluGluAspSerGlySerSerArgLysLysIleThrCysGluCysThrLys 417  
Db 1548 AAATGTACCCAGAAGATTTCAGGTAGCAGGAAGAAAATACATGTGAATGTACTAAA 1607  
Qy 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431  
Db 1608 CCTGATTTTATCCACTTTTCGATGATGTTTCTTCGACGATCC 1649

RESULT 30  
PFAMSP8351  
LOCUS  
DEFINITION P. falciparum DNA for the precursor to the major merozoite surface proteins, C-terminal.  
ACCESSION D13348  
VERSION D13348.1  
KEYWORDS EGF-like domains; major merozoite surface protein precursor.  
SOURCE Plasmodium falciparum (isolate 835/1) merozoite, DNA.  
ORGANISM Plasmodium falciparum

Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
1 (bases 1 to 1726)  
Jongwutiwes, S., Tanabe, K. and Kanbara, H.  
Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP1) of Plasmodium falciparum from field isolates  
Mol. Biochem. Parasitol. 59 (1), 95-100 (1993)  
93295445  
Jongwutiwes, S.  
Direct Submission  
Submitted (05-OCT-1992) Somchai Jongwutiwes, Institute of Tropical Medicine, Nagasaki Univ., Dept. of Protozoology; 12-4 Sakamoto-machi, Nagasaki 852, Japan (Tel:0958-47-2111(ex.3747), Fax:0958-47-6607)  
Submitted (05-OCT-1992) to DDBJ by:  
Somchai Jongwutiwes  
Department of Protozoology  
Institute of Tropical Medicine  
Nagasaki University  
12-4 Sakamoto-machi  
Nagasaki 852  
Japan  
Phone: 0958-47-2111x3747  
Fax: 0958-47-6607.  
Location/Qualifiers  
1. .1726  
/organism="Plasmodium falciparum"  
/db\_xref="taxon:5833"  
/chromosome="9"  
/dev\_stage="merozoite"  
1. .1712  
/genes="MSP1"  
1. .1712  
/gene="MSP1"  
/codon\_start=3  
/product="major merozoite surface protein precursor"  
/protein\_id="BAA02609.1"  
/db\_xref="GI:391803"  
/translation="PLKTLSEVSIQTEDNYANLEKFRALSKIDGKLNLDNHLGKKKLS  
FLSSGLHLITELKEVINKNTGNSPNNKKNVNAALSKYENFLPEAKVTVVTPPO  
PQVTPSPLSVRVSGSGSTKEETQIPTSGSLLELOQVQLQYDDEDDSLVLPFG  
ESEDNEYLDQVVTGEGATSYMDNLSGFENYDVIYLRPLAGVYKSLKKQIKENIIT  
FNLNLDILNSRLKKRYDFLDVLESDMQFHKHSSNEYIIEFSKLLNSQKNTLLKS  
YKYIKESVENDIKFAQEGISYEKVLAKYKDDLESIKVKIKEKEFPSSPTTTPSPA  
KYDEQKESKFLPFLNIETLYNVLNKIDDIYLNILKAKINDCNKDEKDEAHVKITLS  
DLKAIIDKIDLPKNTDFEAIKKLINDTKMLGLLSTGLVQNPNTTISKLIEGK  
FQDLNLSIQHCYVKQCPENSGCFRHLDEREECKCLLNKYQEGDKCVENPNPTCNENN  
GGCDAKACTEEDSGSNKRKITCECTKPDSPYLPFDGIFCSSNPLGLISFLILMLILY  
SFT"  
BASE COUNT 741 a 233 c 245 g 507 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 1,62e-125 Length: 1726  
Score: 1915.00 Matches: 370  
Percent Similarity: 95.94% Conservative: 8  
Best Local Similarity: 93.91% Mismatches: 10  
Query Match: 84.18% Indels: 6  
DB: 3 Gaps: 2  
US-10-057-531A-2 (1-431) x PFAMSP8351 (1-1726)  
Qy 38 ThrAspAspAspLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMet 57  
Db 486 TCCGAAGATAATGACGAATATTTAGATCAAGTACTAGTGCAGAA----- 530  
Qy 58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77  
Db 531 GCAATATCTGTCACATGGATAATATCTCTCAGGATTTGAAAATGAATGATGTTATA 590  
Qy 78 TyrlLeuLysProLeuAlaGlyValTyrArgSerLeuLysGlnIleGluLysAsnIle 97

Db	591	TATTATAAACCTTTAGCTGGAGTATATAGAACCTTAAAAAACAACAAATTCAAAAAACATT	650
Qy	98	PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr	117
Db	651	ATTACATTTAAATTTAAATTTGAACGATATCTTAAATTCAGCTCTTAAGAAACGAAATAT	710
Qy	118	PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr	137
Db	711	TTCTTAGATGATTAGATCTGATTATGCAATTTAATGCAATTTAACAATATATCTCAATGAATAC	770
Qy	138	IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer	157
Db	771	ATTATGAGATTCATTTAAATTTAGTTCAGAACAAAAACACACTTTTAAAGAT	830
Qy	158	TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer	177
Db	831	TACAAATATATAAAGAATCAGTACAAATGATATTAATTTGCACAGGATATAAGT	890
Qy	178	TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValIle	197
Db	891	TATTATGAAAGGTTTACGGAATATATAGGATGATTTAGATCAATTAATAAAGTTATC	950
Qy	198	LysGluGluLysGluLysPheProSerProProThrThrProProSerProAlaLys	217
Db	951	AAAGAAGAAAGAGGAG--TTCCCATCATCACCAACCAACACCTCCGTCACCACCAAA	1007
Qy	218	ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu	237
Db	1008	ACAGACGAAACAAAGAGAAAGTAAGTTCCTTCCTCCATTTTAAACAAACATTGAGACCTTA	1067
Qy	238	TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn	257
Db	1068	TACAATACTTAGTTAATAAATTCAGCATTTACTTAATTAATTAAGCAAGATTAAC	1127
Qy	258	AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys	277
Db	1128	GATTCTAATGTTGAAAGAGATGAACACACATGTTAAATAAATACTTAATCTAGTATTAAA	1187
Qy	278	AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys	297
Db	1188	GCAATTTGATGACAAAATAGATCTTTTAAAAACACATACAGCATTCGAAGCAATTA	1247
Qy	298	LeuIleAsnAspAspThrLysLysAspMetLeuGluLysLeuLeuSerThrGlyLeuVal	317
Db	1248	TTGATAATGATGATACGAAAGAAAGATATGCTTGCAATTAATCTTAGTACAGGATTAGT	1307
Qy	318	GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeu	337
Db	1308	CAAAATTTTCTTAATACATAATATCAAAATTAATTTGAAGGAAATTCACAGATATGTA	1367
Qy	338	AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg	357
Db	1368	AACATTTTCAACACACCAATGCGTAAAAAACAATGTCAGAAAATTCCTGGATGTTTACA	1427
Qy	358	HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys	377
Db	1428	CATTTAGATGAAGAGAGAATGTAATGTTTATTAATTAACAACAAGAGGTGATAAA	1487
Qy	378	CysValGluAsnProAsnProThrCysAsnGlnAsnAsnGlyGlyCysAspAlaAspAla	397
Db	1488	TGTGTTGAAATCCAAATCCTACTTGTAAACGAAATAATGTTGGATGATGTCAGATGCC	1547
Qy	398	ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys	417
Db	1548	AAATGTACCGAAGAGATTCAGGTAGCAACGAAAGAAATACATGTGTAATGTACTAAA	1607
Qy	418	ProAspSerTyrProLeuPheAspGlyIlePheCysSerSer	431
Db	1608	CCTGATCTTATCCACTTTTCATGATGATTTTCTGCAGTTCC	1649
LOCUS	PFAMSP8352 1726 bp DNA linear INV 29-MAY-2002		
RESULT 31	PFAMSP8352		

DEFINITION	P. falciparum DNA for the precursor to the major merozoite surface proteins, C-terminal.		
ACCESSION	D13349		
VERSION	D13349.1 GI:391804		
KEYWORDS	EGF-like domains; major merozoite surface protein precursor.		
SOURCE	Plasmodium falciparum (isolate 835/2) merozoite, DNA.		
ORGANISM	Plasmodium falciparum		
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
AUTHORS	1 (bases 1 to 1726)		
TITLE	Jongwutiwes, S., Tanabe, K. and Kanbara, H.		
JOURNAL	Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP1) of Plasmodium falciparum from field isolates		
REFERENCE	Mol. Biochem. Parasitol. 59 (1), 95-100 (1993)		
AUTHORS	Jongwutiwes, S.		
TITLE	2 (bases 1 to 1726)		
JOURNAL	Direct Submission		
REFERENCE	Submitted (05-OCT-1992) Somchai Jongwutiwes, Institute of Tropical Medicine, Nagasaki Univ., Dept. of Protozoology; 12-4 Sakamoto-machi, Nagasaki 852, Japan (Tel:0958-47-2111(ex.3747), Fax:0958-47-6607)		
AUTHORS	Submitted (05-OCT-1992) to DDBJ by:		
TITLE	Somchai Jongwutiwes		
JOURNAL	Department of Protozoology		
REFERENCE	Institute of Tropical Medicine		
AUTHORS	Nagasaki University		
TITLE	12-4 Sakamoto-machi		
JOURNAL	Nagasaki 852		
COMMENT	Japan		
FEATURES	source		
gene	0958-47-2111x3747		
CDS	0958-47-6607.		
FEATURES	Location/Qualifiers		
gene	1..1726		
CDS	/organism="Plasmodium falciparum"		
FEATURES	/db_xref="taxon:5833"		
gene	/chromosome="9"		
CDS	/dev_stage="merozoite"		
FEATURES	1..1712		
gene	/gene="MSP1"		
CDS	<1..1712		
FEATURES	/gene="MSP1"		
gene	/codon_start=3		
CDS	/product="major merozoite surface protein precursor"		
FEATURES	/protein_id="BAA02610.1"		
gene	/db_xref="GI:391805"		
CDS	/translation="PLKLTSEVSIQTEDNYANLEKFRVLSKIDGKLNLDNLHGGKKLS		
FEATURES	FLSSGLHLITELKEVTKNKTGNSPSENNKYNKALKSVENLPEAKVTVTVPQ		
gene	PDVTPSPLSVRVSGSGSTKEETQIPTSGSLTTELOQVQSONYDEEDDSLVLPFG		
CDS	ESENDEYLDQVVTGEALISVTMDNLSGFENEYDVIYLLPLAGYRSLUKQIEKNIT		
FEATURES	FNLNDILNSRLAKRKYFLDVLSDLMQFKHISSEYIIEDESKLNSKQNTLLKS		
gene	YKYIKESVENDIKFAQEGISYKVLAKYKDDLESIKKVIKEKEFFSPPTPPSPA		
CDS	KTDBQKESKPLPFLTNIETIYLNVLNKIDDIYLNKAKINDCNVEKDEAHVITKLS		
FEATURES	DLKAIIDDKIDLKFNNDPEAIKLIIDDTKKDMLGKLLSTGLVNFNTIISKLEIGK		
gene	FODMLNSHQCVKVKQCPENSGCFRHLDERECKLLNKGKQCKVCVENPNTCNENN		
CDS	GGCDANAKCTEEDSGSGNGKITCTCTPDSPFLDFGIFCSSSNFGISFLILMLLY		
FEATURES	SFI"		
BASE COUNT	741 a	233 c	245 g 507 t
ORIGIN			
Alignment Scores:			
Pred. No.:	1.62e-125	Length:	1726
Score:	1915.00	Matches:	370
Percent Similarity:	95.94%	Conservative:	8
Best Local Similarity:	93.91%	Mismatches:	10
Query Match:	84.18%	Indels:	6
DB:	3	Gaps:	2
US-10-057-531A-2	(1-431) x	PFAMSP8352	(1-1726)
Qy	38	ThrAspAspAspAspLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMet	57
Db	486	TCCGAGATATGACCAATATTTAGATCAAGTAGTAACTGGAGAA-----	530

QY 58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77  
 |||||  
 Db 531 GCAATATCTGTCAACATGGATATATCTCTCAGGATTGGAAATGAATATGATGTATA 590  
 |||||  
 QY 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysGlnIleGluLysAsnIle 97  
 |||||  
 Db 591 TATTTAAACCTTTAGCTGGATATATAGAAAGCTTAAAAACAAATTTGAAAAACAT 650  
 |||||  
 QY 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117  
 |||||  
 Db 651 ATTACATTTAATTTAATTTGAACGATATCTTAAATTCACGCTTAAAGAAACCAATAT 710  
 |||||  
 QY 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137  
 |||||  
 Db 711 TTCTTAGATGATTACAAATCTGATTAAATGCAATTTAAACATATATCTCCAAATGAATAC 770  
 |||||  
 QY 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157  
 |||||  
 Db 771 ATTATTGAAGATTCAATTTAATTTAATTTGAATTCAGAAACAAACACACTTTTAAAAAGT 830  
 |||||  
 QY 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177  
 |||||  
 Db 831 TACAATATATAAAGAAATCAGTAGAAATGATATTAATTTGCACGAAAGGTATTAAGT 890  
 |||||  
 QY 178 TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValIle 197  
 |||||  
 Db 891 TATTATGAAAGGTTTACGAAATATATAGGATGATTAGATCAATTAATAAAGGTTATC 950  
 |||||  
 QY 198 LysGluGluLysGluLysPheProSerSerProProThrThrProProSerProAlaLys 217  
 |||||  
 Db 951 AAAGAAGAAAGAG---TTCCCATCATCACCAACCAACACTCCGTCACCAACGAAAA 1007  
 |||||  
 QY 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237  
 |||||  
 Db 1008 ACAGACGACAAAGAAAGAAAGTAAGTTCCTTCATTTTAAACAACATTTGAGACCTTA 1067  
 |||||  
 QY 238 TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn 257  
 |||||  
 Db 1068 TACAATAACTTAGTTAATAAAATGACGATTACTTAATTAACCTTAAGGCAAGATTAAC 1127  
 |||||  
 QY 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277  
 |||||  
 Db 1128 GATTGTAACTGTGAAAGATGATGACACATGTTAAATAACTAACTTACTGATTTAAAA 1187  
 |||||  
 QY 278 AlaIleAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaLysLysLys 297  
 |||||  
 Db 1188 GCAATTTGATGACAAATATAGATCTTTTAAACACTTAACGACTTCGAACCAATTAATA 1247  
 |||||  
 QY 298 LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal 317  
 |||||  
 Db 1248 TTGATAAATGATGATACGAAAAAGATATGCTTGGCAAAATTACTTAGTACAGGATTAGTT 1307  
 |||||  
 QY 318 GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeu 337  
 |||||  
 Db 1308 CAAAATTTTCCCTAATAACAATAATATCAAAATTAATTTGAAGAAATTTCCAAGATATGTTA 1367  
 |||||  
 QY 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357  
 |||||  
 Db 1368 AACATTTTCAACACCAACCAATGCGTAAAAAACAATATGTCAGAAATTTCTGGATGTTTCAGA 1427  
 |||||  
 QY 358 HisLeuAspGluArgGluGluCysLysCysLysLeuLeuAsnTyrLysGlnGluGlyAspLys 377  
 |||||  
 Db 1428 CATTTAGATGAAAGACAAAGAAATGTAATGTTTATTAATTAACAAACAAAGAGGTGATAA 1487  
 |||||  
 QY 378 CysValGluAsnProAsnProThrCysAsnGluAsnGlyGlyCysAspAlaAspAla 397  
 |||||  
 Db 1488 TGTGTTGAAATCCAAATCCTACTTGAACGAAATTAATGGTGGATGTTGATGACAGATGCC 1547  
 |||||  
 QY 398 ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417  
 |||||  
 Db 1548 AATGTACCCGAAGAGATTTCAGTAGCAACGGAAGAAATACATGTTGAATGTACTATAA 1607  
 |||||

QY 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431  
 |||||  
 Db 1608 CCTGATTCATCCACTTTTCGATGATGATTTCTTCGAGTTCC 1649  
 |||||  
 RESULT 32  
 PFAMSP838  
 LOCUS  
 DEFINITION  
 P. falciparum DNA for the precursor to the major merozoite surface proteins, C-terminal.  
 1726 bp DNA linear INV 29-MAY-2002  
 ACCESSION  
 D13350  
 VERSION  
 D13350.1 GI:391808  
 EGF-like domains; major merozoite surface protein precursor.  
 SOURCE  
 Plasmodium falciparum (isolate 838) merozoite, DNA.  
 ORGANISM  
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 REFERENCE  
 1 (bases 1 to 1726)  
 Jongsutwittes,S., Tanabe,K. and Kanbara,H.  
 Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP1) of Plasmodium falciparum from field isolates  
 TITLE  
 JOURNAL  
 MEDLINE  
 93295445  
 REFERENCE  
 2 (bases 1 to 1726)  
 Jongsutwittes,S.  
 Direct Submission  
 Submitted (05-OCT-1992) Somchai Jongsutwittes, Institute of Tropical Medicine, Nagasaki Univ., Dept. of Protozoology; 12-4 Sakamoto-machi, Nagasaki 852, Japan (Tel:0958-47-2111(ex.3747), Fax:0958-47-6607)  
 COMMENT  
 Submitted (05-OCT-1992) to DDBJ by:  
 Somchai Jongsutwittes  
 Department of Protozoology  
 Institute of Tropical Medicine  
 Nagasaki University  
 12-4 Sakamoto-machi  
 Nagasaki 852  
 Japan  
 Phone: 0958-47-2111x3747  
 Fax: 0958-47-6607.  
 FEATURES  
 Location/Qualifiers  
 1..1726  
 /organism="Plasmodium falciparum"  
 /db\_xref="taxon:5833"  
 /chromosome="9"  
 /dev\_stage="merozoite"  
 1..1712  
 /gene="MSP1"  
 <1..1712  
 /gene="MSP1"  
 /codon\_start=3  
 /product="major merozoite surface protein precursor"  
 /protein\_id="BAA02611.1"  
 /db\_xref="GI:391809"  
 /translation="PLKTLSEVSIOTEDYANLEKPRVLKSDIGKLNHLHGKKKLS  
 FUSGSLHLHITELKEVINKNKNTGNSPNNKVAEALSKSYENFLPEAKVTVPFPQ  
 PVTPLSLSVRSVSGSGSTKEETQTSGLLTQLQVQVQLQYDEDSLVLPFG  
 ESEDNEDVLDQVVTGEATSVTMDNLSGFENEVDVYLKPLAGVYRLSKOIEKNIT  
 FNLNLDLIILNSRLKRYKFLVDLSLQDFKHISSEYIIEDSKLLNSEOKNTLLKS  
 YKIKESVENDIKFAQEGISYEKVLAKYKDDLESIKKVIKEKEFPSSPPTPPSPA  
 KYDEQKESKFLPFLNTIETLNLYNKNIDDDYLINLAKINDCNVKEKDEAHKRTKLS  
 DLKATDDKIDLEKNTNDFAIKKLINDDTKDMGLKLLSTGLVQFPNTIISKLEGG  
 FODMLNISQHCVRKQCPENSGCFRHLDERECKLLNYKQEGDKCVENPNPTCENN  
 GGCDADAKCTEEDSGSGNKKITCECTKPDSPYLPFDGIFCSSSNFLGIFLLILMLLY  
 SPI"  
 BASE COUNT 741 a 232 c 245 g 508 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1-62e-125 Length: 1726  
 Score: 1915.00 Matches: 370  
 Percent Similarity: 95.94% Conservative: 8  
 Best Local Similarity: 93.91% Mismatches: 10  
 Query Match: 84.18% Indels: 6





Db	1368	ACACATTTTCACACACACCAATGCGTAAAAAACAATGTCACAGAAAATTCCTGGATGTTTCAGA	1427
Qy	358	HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAsPlys	377
Db	1428	CATTTCAGATCAAGAGAGAAGATGTAATGTTTATTAAATACAAACAAGAAGGTGATAAA	1487
Qy	378	CysValGluAsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAla	397
Db	1488	TGTGTTCAAAATCCAAATCCTACTGTTAAGCAAAATTAATGGTGGATGTGATGTCAGATGCC	1547
Qy	398	ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys	417
Db	1548	AAATGTACCGAGAAGATTCAGGTAGCAACGGAAGAAATACATGTGATGTACTATAA	1607
Qy	418	ProAspSerTyrProLeuPheAspGlyIlePheCysSerSer	431
Db	1608	CCTGATCTTATCCACTTTTCGATGGTATTTCCTGCAGTTCC	1649
RESULT	34		
LOCUS	AF218248	5085 bp DNA linear	INV 24-JAN-2000
DEFINITION	Plasmodium falciparum major merozoite surface antigen (gp195) gene, complete cds.		
ACCESSION	AF218248		
KEYWORDS	AF218248.1	GI:6740099	
SOURCE	Plasmodium falciparum.		
ORGANISM	Plasmodium falciparum.		
REFERENCES	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
AUTHORS	1 (bases 1 to 5085)		
TITLE	Shan,Z.X., Yu,X.B., Li,X.R., Ma,C.L., and Fang,J.M.		
JOURNAL	Molecular cloning and sequence analysis of major merozoite surface antigen(gp195)gene of Plasmodium falciparum isolate FCCI/HN		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 5085)		
TITLE	Shan,Z.X.		
JOURNAL	Direct Submission		
FEATURES	Submitted (21-DEC-1999), Department of Parasitology, Sun Yat-sen University of Medical Sciences, 74 Zhongshan Road II, Guangzhou, Guangdong 510089, P.R.China		
source	Location/Qualifiers		
gene	1..5085		
CDS	/organism="Plasmodium falciparum"		
	/isolate="FCCI/HN"		
	/db_xref="taxon:5833"		
	/country="China"		
	1..5085		
	/gene="gp195"		
	1..5085		
	/codon_start=1		
	/product="major merozoite surface antigen"		
	/protein_id="AAF27526.1"		
	/db_xref="GI:6740100"		
	/translation="MKTIFFLCSFLPFIINTCTHESYOLVKLLEALEDAVLGTYS		
	LFQEKWLNQEGTGAATVTTSTPGSGSVTSGGGSVASVSGSGSGSGSGSGS		
	RTNPNSNSDPAKSVADKLHRVQNLFTIKELYPELDLNNHMLTCDNTHGFKY		
	LDIGYEINELLVKNLYIDKLRALCANASYCQIPFNLRKINRANEDLVKLTIVGY		
	RKPLDNKDNVNGMKYIKNNKTTIANINELIGSKRTIDQNKADNEEGKKLYQQA		
	YNFIYKQIQEAAHNLITSEKRLIDKLNENIKKLEDIDKTKDAEKLTGTSKPNP		
	LPNNKQQLVSTTQPLVAASSETTEDGSHSTPTLSQSGTEETEEVTVGTTVTITL		
	LPNNKQQLVSTTQPLVAASSETTEDGSHSTPTLSQSGTEETEEVTVGTTVTITL		
	KSVQIPKVPNGIVPLPTDTHNSLAADNDKNSYGLMNPDTKEINKIITDNKE		
	RKIFINNIKKQIDLEEKNIHTKEQNKLLDEYKSKKDEEELLEKEYEMKFNPNP		
	DVVDKIPNTINVEKQKNKPFSSNNSYNVQKKKALSLEYDLSLRKGEISEKOF		
	NHYYITLKGLEADFLKKALETEIKSENKLEIKNFKGLTHSANASLEVDIKVLQVQK		
	LKKIEDRLKIEFLKKALEKDSHTVPIYKPNKPEPYIIVLKKEVDKLEIKPW		
	KDMKKEQAVLSSTQPLVAASSETTEDGSHSTPTLSQSGTEETEEVTVGTTVTITL		
	PPTQSPPKPKVVENSIKHSNDNSOALTQTVYLLKLDLEFTKSYICHKYLIVNSSS		
	MDQKLLVYNLTPEEKNELSCDPLDLFNQNNIPAMSYLSDMNNDLQHLFELVQ		
	KEMIIYLHKLKEENHIKKLEEQKQITGTSTSGPGNTVNNQAQTHSNQSQNSQA		
	SSYNTQNGAVSSGPAVESEHDPLTVLSLSDNLQIVLNNKTKVNPPLTISTT		
	EMEKFYENILKNDITFYNDIKIIFQVKKNSKVTIGLTETQKNALNDEIKKLKDTQLS		
	DLVNYKKLLDRLFNNKKELGQDKMKKSLTLLKEQLSKLNSLNNPHNVLQNFVSF		

Alignment Scores:  
Pred. No.: 5, 04e-125 Length: 5085  
Score: 1915.00 Matches: 370  
Percent Similarity: 95.94% Conservative: 8  
Best Local Similarity: 93.91% Mismatches: 10  
Query Match: 84.18% Indels: 6  
DB: 3 Gaps: 2

US-10-057-531A-2 (1-431) x AF218248 (1-5085)

BASE COUNT 2244 a 671 c 678 g 1492 t  
ORIGIN

KKKEAEIAETENTLENTKILLKHGKLVKYYNGESSPLKTLSEVSIQTDNYANLEK  
FRVLSIDGKLNLDNLHKKKFLSSGLHLHTELKEVINKNTYNGSPSENKKV  
EALKSYLPEAKVTTPVPOPDPSPSVRSVSGSGSTKEETQIPTSGSLTEL  
QOVQSONDEEDSLVLP1RGESEDNDEYLDVVGTGAISVMTDMNLSGFENEYDV  
IYLPLGAVRSLLKQLEKILITFNLANLILNSRLKK8KYFLDVLDESLOMOPKHLS  
NEYIIEDESKLLNSEQKNLLKSKYKIKESVENDIKFAOEGISYKVKVLAAYKDDLES  
IKKVIKEEFSSPTTPPTSPATDEQKESKFLPFLNIETLYNNLVNKIDDDYLIN  
LRAKINDCNVEKDEAHVITKLSDLKAIIDDKIDLFKNTNDFEAKKLIINDTFKOMLG  
KLISLGVQNFQNPNTIISKLEIGKFOOMLNIHQCVKQKOCPENSGCFRHLDRERECK  
LLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSPYLPD  
GIFCSSENFELGISFLILMLILYSFI"

38 ThrAspAspAspLysLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMet 57  
Db 3859 TCCGAAGATAATGACGAATATTAGATCAAGTAACTAAGTAACTGACGAA----- 3903

58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77  
Db 3904 GCAATATCTGTCAAGATGGATAATATCTCTCAGGATTTGAAATGAATATGATGTTATA 3963

78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97  
Db 3964 TATTAAACCTTTAGCTGGATATATAGAAGCTTAAAAACAAATTTGAAAAACATT 4023

98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117  
Db 4024 ATTACATTTTAAATTTTGAACGATATCTTAAATTCAGCTCTTAAAGAACGAAAAATAT 4083

118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137  
Db 4084 TTCTTAGATGTATTAGATCTGATTTAATGCAATTTTAAACATATATCTCAATGAATAC 4143

138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157  
Db 4144 ATTATCAAGATTTCATTTAAATTTATTCAGATTCAGACAAACAAACACACTTTTAAAAAGT 4203

158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177  
Db 4204 TACAAATATATAAAGAAATCAGTAGAAAATGATATAATTTTGCACAGGAGGTAAAGT 4263

178 TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysValIle 197  
Db 4264 TATTATGAAAAGTTTATAGCAATATAAGGATGATTAGATCAATTAATAAAGTTATC 4323

198 LysGluGluLysGluLysPheProSerSerProProThrThrProProSerProAlaLys 217  
Db 4324 AAGAAGAAAAAGGAG---TTCCCATCATCACCACCAACACACCTCGCTCACCAGCAAAA 4380

218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237  
Db 4381 ACAGCAGCAACAAAAAGAGGAAGTAAGTTCCTCCATTTTAAACAAACATTGAGACCTTA 4440

238 TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn 257  
Db 4441 TACAATAACTTAGTTATAAATTTAGCATTTACTTAAATTTAAAGGCAAGATTAAAC 4500

258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277  
Db 4501 GATTGTAAATGTTGAAAAAGATGAAGCAGCATGTTAAAAATACTAACTAGTATTGATTTAAAA 4560

278 AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys 297

Db 4561 GCAATTGATCACAAAATAGATCTCTTTTAAAAACACTAACGACTTCGAAGCAATTAATAAAA 4620

Qy 298 LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal 317

Db 4621 TTGATAAATGATGATACGAAAAAGATATCTCTGGCAAAATTTACTTAGTACAGGATTAGTT 4680

Qy 318 GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeu 337

Db 4681 CAAAATTTTCTTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 4740

Qy 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357

Db 4741 AACATTTTCAACACCAATCGTAAAAAAAACAAATGTCCAGAAAAATTCGTGATGTTTCAGA 4800

Qy 358 HisLeuAspGluArgGluGluCysLysCysLeuAsnTyrLysGlnGluGlyAspLys 377

Db 4801 CATTTAGTAGAAGAGAAAGTAAATGTTTAAATTTAAATTTAAATTTAAATTTAAATTTAA 4860

Qy 378 CysValGluAsnProAsnProThrCysAsnGluAsnGlyGlyCysAspAlaAspAla 397

Db 4861 TGTGTTGAAAAATCAAAATCTACTTGTACGAAAAATAATGTTGATGATGATGATGCC 4920

Qy 398 ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417

Db 4921 AAATGTACCCGAAGAAGATTGAGTAGACACGAGAAAGAAATACATGTGAATGTACTAAA 4980

Qy 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431

Db 4981 CCGATTCCTTATCATCCTTTTCGATGATGATGATGATGATGATGATGATGATGATGATG 5022

RESULT 35  
AF062348 5243 bp DNA linear INV 09-FEB-2001  
LOCUS Plasmodium falciparum strain HNI merozoite surface protein 1  
DEFINITION precursor (msp1) gene, complete cds.  
ACCESSION AF062348  
VERSION AF062348.1 GI:3859842  
KEYWORDS Plasmodium falciparum.  
SOURCE Plasmodium falciparum.  
ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
REFERENCE 1 (bases 1 to 5243)  
AUTHORS Jiang,G., Liu,R., Daubenberger,C.A. and Pluschke,G.  
TITLE Sequence analysis of the MSP 1 gene of Plasmodium falciparum from Hainan, China  
JOURNAL Chung Kuo Chi Sheng Chung Hsueh Yu Chi Sheng Chung Ping Tsa Chih 17 (5), 294-297 (1999)  
REFERENCE 2 (bases 1 to 5243)  
AUTHORS Jiang,G., Liu,R.Z., Daubenberger,C.A. and Pluschke,G.  
TITLE Direct Submission  
JOURNAL Submitted (30-APR-1998) Molecular Immunology, Swiss Tropical Institute, Socinstrasse 57, Basel CH-4002, Switzerland  
FEATURES  
Location/Qualifiers  
1..5243  
/organism="Plasmodium falciparum"  
/strain="HNI"  
/db\_xref="taxon:5833"  
/chromosome="9"  
/country="China"  
<66..>5150  
/gene="msp1"  
/gene="msp1"  
/product="merozoite surface protein 1 precursor"  
66..5150  
/gene="msp1"  
/note="binds spectrin"  
/codon\_start=1  
/product="merozoite surface protein 1 precursor"  
/protein\_id="AAC72884.1"  
/db\_xref="GI:3859843"  
/translation="MKLIIFLCSFLFIINTQCVTHTESYQELVKKLEALEDVLTGYS"

Alignment Scores:	5.2e-125	Length:	5243
Score:	1915.00	Matches:	370
Percent Similarity:	95.94%	Conservative:	10
Best Local Similarity:	93.91%	Mismatches:	8
Query Match:	84.18%	Indels:	6
DB:	3	Gaps:	2
US-10-057-531A-2 (1-431) * AF062348 (1-5243)			
QY	38	ThrAspAspAspLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMet	57
DB	3924	TCCGAAGATAATGACGAATAATTAGATCAAGTAGTAGTACTGGAGAA-----	3968
QY	58	AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle	77
DB	3969	GCAATATCTCTCACAAATGGGATATATCTCTCAGGATTTTGAAATGAATATGATGTATA	4028
QY	78	TyrLeuLysproLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle	97
DB	4029	TATTTTAAACCTTTTACCTGGAGTATATAGAAGCTTAAAAAACAATTTGAAAAACAAT	4088
QY	98	PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr	117
DB	4089	ATTACATTTTAATTTAAATTTGAACGATATCTTAAATTCACGCTCTAAGAAACGAATAT	4148
QY	118	PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr	137
DB	4149	TTCTTTAGATGTATTAGAAATCTGATTTTAAATGCAATTTTAAACATATATCCTCAAAATGAATAC	4208
QY	138	IleIleGluAspSerPheLysLeuLeuAsnSerGluClnLysAsnThrLeuLeuLysSer	157
DB	4209	ATTATTGAAGATTCAVTTTAAATTTATTGAATTCAGAACAAAAACACACTTTTAAAAAGT	4268
QY	158	TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer	177
DB	4269	TACAATATATAAAGAAATCAGTAGAAAATGATATTAATTTGCACAGGAAGGTATTAAGT	4328
QY	178	TyrTyrGluLysValLeuAlaTyrLysAspAspLeuGluSerIleLysLysValIle	197
DB	4329	TATTTATGAAAGGTTTTAGCGAAATATAGGATGTATTTAGAAATCAATTTAAAAAGTTATC	4388

/organism="Plasmodium falciparum"  
 /strain="HN2"  
 /db\_xref="taxon:5833"  
 /chromosome="9"  
 /country="China"  
 <66..>5180  
 /gene="msp1"  
 <66..>5180  
 /gene="msp1"  
 /product="merozoite surface protein 1 precursor"  
 66..5180  
 /gene="msp1"  
 /note="binds spectrin"  
 /codon\_start=1  
 /product="merozoite surface protein 1 precursor"  
 /protein\_id="AAC72885.1"  
 /db\_xref="GI:3859845"  
 /translation="MKTIIFLCFLPFIINTOCVTHESYOELVKKLEALEDAVLGYS  
 LFOKEKRVNLTSPNSDSDAKSYADUKHRVONYLFTIKELKYPELEDTLTHMLTLCND  
 IHGFKYLIDGEEINELLYKLNFYDLRAKLDACANSYCOIPFNKIRANELDVLK  
 KIVFGYKPLDNIDKNVGMEDYIKKNKTTIANINELIGSKKTIIDONKNADNEGGK  
 KLYAOVNLFIYNKLOEAHNLISVLEKRTDILKKNENIKKLLDKINEIKNPPANSK  
 NPTNLLDKKLEIEEHEEKIEIAKTIKFNIDSLFTDPLELEYLREKKNKVDVTPGS  
 QDPTKSVQIPKVPNGIYVPLPLTDHNSLAADNKNSTGDLMDNPTREKINEKIT  
 DNKERKIFINIKKQIDLEKNINHTKQNKLEDEYKSKDYEEELERFYEMKFN  
 NFDKVDVKIKTSARYTYNVEKRNKNSFNSSNVNVOKKKALSYLEDYSLRKGIS  
 EKDFNHYTLTKLEADIKLLEIKSENKILEKNFKGLTHSANASLEVSIVKLQV  
 QVILLIKLEIDKLEIFLKNALQKDSIHVPNIYKPNKPEPYLLVLKKEVDKLEF  
 IPKVKMLKKEQAVLSITOPVAASETTEDGHSHTLSOGSETVETETETETV  
 GHVTVTITLPTQPPKPKVENSIEHKSNDNSOALTKTVYLKDLDFLKSytic  
 HKTILVNSNDNLTPEERNEKSLCDPLDLFIQNNIPAMYSLYDSMND  
 LQHLFFELYQKEMYYLHKLKEENHKKLEEQKQITGSTSPGNTVNTAOSATH  
 SNGQNSASNTONGVAVSSGPAAVEEDHPDLVLSINDLKGIVSLNLGNKTK  
 VNPATISTTEMEKFEYENILKNNDTVFNDDIKQFVKSNSKVITGLTETOKNALNDEIK  
 KLDTQLSLDKYKVLKDLRFNKKELGQDMOIKKTLILKLEOLESLNKNPH  
 NVLQNSVFENKKEAEIAENTLNTKILLKHYGLVYKYNIGESPPLTKLEVSIO  
 TEDNYANLEKFRVLSIDGNLNLHLGKKLFLSGLSHLLEKVEITKNKNTGN  
 SPSENKNEAKLSVFLPEARVTVTPQPDVTPSPLSVRSVSGSGSTKEETQI  
 PTSGSLTELQVQVQSONYDEEDLSQVLEIFGESEDNDEYLDQVTKGEAISYVLEI  
 LSGFENEYDIYLPKAGVRSLLKQTEKNIITFNLDNLINSRLKRYKRYFDVLES  
 DLMQFHISSEYIIEDSFKLNSQKNTLLKSYKIKESVENDIKFAQBGISYKVK  
 LAKYKDLSEKVIKVEKEFPSPPTTPSPAKTDEQKESKFLPLNTIETLYNNL  
 VNKIDYLYNLKAKINDCNVEKDEAHVKITKLSDLKAIDDKLFXNTDFEAKKLI  
 NDLTKDMGLKSLGVLQNFNPTIISKLEIGKFPQDMNISQHCVKKQCPENSGCFR  
 HDBRECKCLLYNKQEGDKVCENPNTENNGGCDADAKTEEDSGSGNKKITCEC  
 TKPDSPVLFQIFCSSNFIISFLILMLILYSFI"

gene

mRNA

CDS

Db 4119 ATTACATTAAATTTAAATTTTCAACGATATCTTAAATTCACGCTCTTAAGAAACGAAATAT 4178  
 Qy 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137  
 Db 4179 TTCCTAGATGATTAGAAATCTGATTAATCAATTTAAACATATATCTCAATGAATAC 4238  
 Qy 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157  
 Db 4239 ATTATTGAAGATTTCATTTAAATTTGAATTCAGAACAAAAACACACTTTTAAAGT 4298  
 Qy 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluIleSer 177  
 Db 4299 TACAATATATAAAGAAATCAGTAAATATTAATTTGACACAGAAAGGTATAAGT 4358  
 Qy 178 TyrTyrGluLysValLeuAlaLysTyrLysAspLeuGluSerIleLysLysValIle 197  
 Db 4359 TATTATGAAAGGTTTGTAGCGAAATTAAGGATGATTTAGAAATCAATTTAAAGGTATTC 4418  
 Qy 198 LysGluGluLysGluLysPheProSerSerProProThrThrProProSerProAlaLys 217  
 Db 4419 AAAGAAGAAAAAGGAG---TTCCCATCATCACCCACCAACACACCTCCGTCACCAACAAA 4475  
 Qy 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237  
 Db 4476 ACAGACCAACAAAAAGAGAAAGTAAGTTCCTCCATTTTAAACAAACATTTGAGACCTTA 4535  
 Qy 238 TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn 257  
 Db 4536 TACAAATACCTTAGTAAATAAATTCAGATTACCTTAATTAATTAAGGCAAGATTAAC 4595  
 Qy 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277  
 Db 4596 GATTGTAATGTTGAAAGATGAAGACATCTTAAATAAATACTAACTAGTATTTAAAA 4655  
 Qy 278 AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys 297  
 Db 4656 GCAATTGATGACAAAATAGATCTTTTAAAAAACAACACTAACGACTTCGAAGCAATTTAAAAA 4715  
 Qy 298 LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal 317  
 Db 4716 TTGATAATGATGATACGAAAAAGATATGCTGGCAATTTACTTAGTACAGGATAGTT 4775  
 Qy 318 GlnAsnPheProAsnThrIleIleSerLysLeuIleGluLysPheGlnAspMetLeu 337  
 Db 4776 CAAAAATTTTCTTAATAACAATAATCAAAATTAATTAAGAGAAAAATTCGAAGATATGTTA 4835  
 Qy 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357  
 Db 4836 AACATTTCAACACCAATCGTAAAAAACAATGTCCAGAAAAATTCGGATGTTTCAGA 4895  
 Qy 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377  
 Db 4896 CATTTAGATGAAGAGAGAAATGTAATGTTTATTAATAACAACAAGAGGTGATAAA 4955  
 Qy 378 CysValGluAsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAla 397  
 Db 4956 TGTGTTGAAATCCAAATCTACTTGTAAACGAAAAATTAATGTTGATGTCAGATGCC 5015  
 Qy 398 ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417  
 Db 5016 AATGTACCCGAAGAAGATTTCAGGTAGCAACGGAAGAAAAATCACATGTGAATGACTAAA 5075  
 Qy 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431  
 Db 5076 CCTGATCTTATCCACTTCCTTCGATGGTATTTCTTCGAGTTCC 5117

RESULT 37

PFU20730

LOCUS

DEFINITION

PFU20730

ACCESSION

1116 bp DNA linear INV 14-FEB-1995  
 Plasmodium falciparum clone B16A merozoite surface protein 1  
 (MSP-1) gene, 42 kDa C-terminal region, partial cds.

## Alignment Scores:

Pred. No.: 5,23e-125 Length: 5271  
 Score: 1915.00 Matches: 370  
 Percent Similarity: 95.94% Conservative: 8  
 Best Local Similarity: 93.91% Mismatches: 10  
 Query Match: 84.18% Indels: 6  
 DB: 3 Gaps: 2

US-10-057-531A-2 (1-431) x AF062349 (1-5271)

Qy 38 ThrAspAspAspLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMet 57  
 Db 3954 TCCCAAGATATGACGATATTTAGATCAAGTAGTACTGAGAA----- 3998  
 Qy 58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77  
 Db 3999 GCAATATCTGCAACATGATAATATCCTCTCAGGATTTGAAATGAATGATGTTATA 4058  
 Qy 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97  
 Db 4059 TATTAAACCTTTAGCTGGAGTATATAGAACCTTAAACAAACAAATTTGAAAAACATT 4118  
 Qy 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLeuLysLysArgLysTyr 117

```

VERSION      U20730.1  GI:564861
SOURCE       malarial parasite.
ORGANISM     Plasmodium falciparum
REFERENCE    Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS      1 (bases 1 to 1116)
              Shi.Y.-P., Alpers.M.P., Povoa.M.M., Nahlen,B.L., Oloo.A.G. and
              Lai.A.A.
TITLE        Sequence of the C-terminal region of merozoite surface protein-1
              (MSP-1) in field-derived Plasmodium falciparum
JOURNAL      Unpublished
AUTHORS      Shi.Y.-P.
TITLE        Direct Submission
JOURNAL      Submitted (07-FEB-1995) Ya-Ping Shi, DPD/CDC 4770, Molecular
              Vaccine Section, Bldg 22, Rm. 8F12, Buford Highway, Chamblee, GA,
              30341 USA
FEATURES     Location/Qualifiers
              1..1116
                /organism="Plasmodium falciparum"
                /isolate="Infected human blood sample from Papua New
                Guinea"
                /db_xref="taxon:5833"
                /clone="B16A"
              gene      1..1116
                /gene="MSP-1"
                <1..>1116
                /gene="MSP-1"
                /codon_start=1
                /product="merozoite surface protein 1, 42 kDa C-terminal
                region"
                /protein_id="AA62221.1"
                /db_xref="GI:564862"
                /translation="AISVTMDNLGSEFENEYDVIVLKLPLAGVRSILKKQIEKNITFN
                LNLNDILNSRLKRVFLVDVLEDMOPKHISSEVIEDSKLLNSEOKNTLLSKYK
                YTKESVDNIEKAQEGISYXEVLAQYKDDLESIKVIEKEFPSPPTTPSPAKT
                DQKKESFELPNIETRIENLVNKKIDYILNKLAKINDCWNERNHAKVTKLSDL
                KAIIDDLQFNHNDFEAIKLINDTTRKDMGLKLLSTGLGVNFPNTIISKLEIKFQ
                NDLNTSQHQQKCPENSGCFRHLDEREECKLLNKOEGDKCVENPNPTCNENNGG
                CDADAKCTEEDSGNGKIKTCECTKPDSPYLPFDGIFCS"
              BASE COUNT      486 a 145 c 160 g 325 t
              ORIGIN
Alignment Scores:
Pred. No.:      4.05e-125      Length:      1116
Score:          1906.50      Matches:      366
Percent Similarity: 98.66%      Conservative: 2
Best Local Similarity: 98.12%      Mismatches: 4
Query Match:      83.80%      Indels:      1
DB:              3          Gaps:      1
US-10-057-531A-2 (1-431) x PFU20730 (1-1116)
QY  58 AlaileSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77
    |||||
Db  1 GCAATATCTGTCACATGGATATATCTCTCAGGATTTGAAATGAATATGATGTTATA 60
    |||||
QY  78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97
    |||||
Db  61 TATTAAACCTTTAGCTGAGTATATAGAGCTTAAAAACAAATTTGAAAAACATT 120
    |||||
QY  98 PheThrPheAsnLeuAsnLeuAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117
    |||||
Db  121 TTATACATTTAATTTGAACGATATCTTAAATTCAGTCTTAAAGAACCAAAATAT 180
    |||||
QY  118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137
    |||||
Db  181 TTCTTAGATGCTATAGAACTGATTAATGCAATTTAAACATATATATCTCAATGATAC 240
    |||||
QY  138 IleileGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLysSer 157
    |||||
Db  241 ATTATTGAAGATTCAATTAATTTAATTTGAATTCAGAACAAAAACACACTTTTAAAAAGT 300
    |||||
QY  158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177
    |||||
Db  301 TACAAATATATAAAGAATCAGTAGAAAATGATATTAAATTTGCACAGGAGGTATAAGT 360
    |||||
QY  178 TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysValIle 197
    |||||
Db  361 TATTATCAAAAGGTTTATAGCGAATATATAGGATGATTAGATCAATTAATAAAGTTATC 420
    |||||
QY  198 LysGluGluLysGluLysPheProSerSerProThrThrProProSerProAlaLys 217
    |||||
Db  421 AAAGAAGAAAAAGGAG--TTTCCCATCATCACCAACAACACCTCGTCACCAGCAAA 477
    |||||
QY  218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237
    |||||
Db  478 ACAGACCAACAAAAAGAGAAAGTAAGTTCTCTTCCATTTTAAACAAACATTTGAGACCTTA 537
    |||||
QY  238 TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn 257
    |||||
Db  538 TACAATAACTTAGTTAATAAATTGACGATTACTTAATTAACTTAAAGCGCAAGATTAAAC 597
    |||||
QY  258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277
    |||||
Db  598 GATTGTAATGTTGAAAAAATGAAGCACATGTTAAAAATACTAAACTTACTGATTTAAAA 657
    |||||
QY  278 AlaileaspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys 297
    |||||
Db  658 GCAATTTGATGACAAATAGATCTTTTAAAAACCATACGACTTCGAAGCAATTTAAAAA 717
    |||||
QY  298 LeuileAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal 317
    |||||
Db  718 TTGATTAATGATGATGACGAAAAAGATATGCTTGGCAAAATTAATTAATTAACGAGGATTAGT 777
    |||||
QY  318 GlnAsnPheProAsnThrIleLeuSerLysLeuIleGluGlyLysPheGlnAspMetLeu 337
    |||||
Db  778 CAAATTTTCTTAATCAATATATCAAAATTAATTAAGAGAAAAATTTCCAAGATATCTTA 837
    |||||
QY  338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357
    |||||
Db  838 AACATTTCAACACCAATCGTAAAAAACAATGTCCAGAAAAATTTCTGGATGTTTTCAGA 897
    |||||
QY  358 HisLeuAspGluArgGluGluCysLysCysLeuAsnTyrLysGlnGluGlyAspLys 377
    |||||
Db  898 CATTTAGATGAAAGAGAAATGTAATGTTTATTAATTAACAAAGAGGTTGATATA 957
    |||||
QY  378 CysValGluAsnProAsnProThrCysAsnGluAsnGlnGlyCysAspAlaAspAla 397
    |||||
Db  958 TGTGTTGAAATCCAAATCCTACTTGTACGAAAAATTAATGTTGATGTCAGATGCC 1017
    |||||
QY  398 ThrCysThrGluGluAspSerGlySerSerArgLysLysIleThrCysGluCysThrLys 417
    |||||
Db  1018 AAATGTACCGAAGAAAGATTCCAGGTAGCAACGAAAGAAAAATCACATGTGAATGTACTATA 1077
    |||||
QY  418 ProAspSerTyrProLeuPheAspGlyIlePheCysSer 430
    |||||
Db  1078 CCTGATTTTATCCACTTTTTCGATGTTATTTTCTGCAAT 1116
    |||||
RESULT 38
PFU20731
LOCUS    1116 bp DNA linear INV 14-FEB-1995
DEFINITION
(MSP-1) gene, 42 kDa C-terminal region, partial cds.
ACCESSION
U20731
VERSION
U20731.1
GI:564863
SOURCE
malarial parasite.
ORGANISM
Plasmodium falciparum
REFERENCE
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS
Shi.Y.-P., Alpers.M.P., Povoa.M.M., Nahlen,B.L., Oloo.A.G. and
Lai.A.A.
TITLE
Sequence of the C-terminal region of merozoite surface protein-1
(MSP-1) in field-derived Plasmodium falciparum
JOURNAL
Unpublished

```

```

REFERENCE 2 (bases 1 to 1116)
AUTHORS Shi,Y.-P.
TITLE Direct Submission
JOURNAL Submitted (07-FEB-1995) Ya-Ping Shi, DPD/CDC 4770, Molecular
Vaccine Section, Bldg 22, Rm. 8F12, Buford Highway, Chamblee, GA,
30341 USA
FEATURES
source Location/Qualifiers
1..1116
/organism="Plasmodium falciparum"
/isolate="infected human blood sample from Papua New
Guinea"
/db_xref="taxon:5833"
/clone="B16C"
1..1116
/gene="MSP-1"
<1..>1116
/gene="MSP-1"
/codon_start=1
/product="merozoite surface protein 1, 42 kDa C-terminal
region"
/db_xref="GI:664864"
/translation="AISVTMDHILSGFENEVDVYLKPLAGVYRSLKKQIEKNIFTN
LNLDILNSRLKRYKYLVDLSEKLMQKHLSSNEYIIESEKLLNSEOKNTLLSKYK
YIKSEVNDIKPAQEGISYEKVLAKYKDDLESKKVKEKEFPSPPTPPSPAKT
DEQKESKFLPFLNHDIEATIKLIINDTKDKMLKLLSTGLVQNPNTIISKLEGRFQ
KADDKIDLFKNHNDIEATIKLIINDTKDKMLKLLSTGLVQNPNTIISKLEGRFQ
DMLNISQHCVKQCPENSGCFRHLDEECCKLLNYKQEGDKCVENPNPTCENNGG
CDADAKCTEEDSGNGKKTCTCECTKPDSPFLFDGIFCS"
BASE COUNT 485 a 146 c 160 g 325 t
ORIGIN
Alignment Scores:
Pred. No.: 9,07e-125 Length: 1116
Score: 1901.50 Matches: 365
Percent Similarity: 98.66% Conservative: 3
Best Local Similarity: 97.86% Mismatches: 4
Query Match: 83.58% Indels: 1
DB: 3 Gaps: 1
US-10-057-531a-2 (1-431) x PFU20731 (1-1116)
Qy 58 AlarileSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77
Db 1 GCATATCTGTCACAAATGATCATATCTCTCAGGATTGTGAAATGAATATGATGTTATA 60
Qy 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97
Db 61 TATTTAAACCTTTAGCTGGAGTATATAGAACCTTAAAAAACAATTTGAAAAACATT 120
Qy 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117
Db 121 TTACATTTAATTTAAATTTGAACGATATCTTAAATTCACGCTCTTAAGAACGAAATAT 180
Qy 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137
Db 181 TTTCTAGATGTTAGATCTGATTTAATGCAATTTAAACATATATCTCAATGAATAC 240
Qy 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLysSer 157
Db 241 ATTATTGAAGATTCATTTAATTTAGATTGCAACACAAAAAACAACACTTTTAAAAAGT 300
Qy 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluLysSer 177
Db 301 TACAAATATATAAAGAATCAGTACAAAATGATATTAATTTGCACAGGAAGGTATAAGT 360
Qy 178 TyrTyrGluLysValLeuAlaLysTyrLysAspLeuGluSerIleLysValIle 197
Db 361 TATTATGAAGAGGTTTACGCAATATAGGATGATTTAGATCAATTTAAAAAGTTATC 420
Qy 198 LysGluGluLysPheProSerProThrThrProThrProThrProThrProThrProThr 217
Db 421 AAAGAAGAAAGGAG---TTCCCATCATCACCACCAACACACCTCCCTCACCAGCAAAA 477

```

```

Qy 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237
Db 478 ACAGACGACAAACAAAGAGAAAGTAAGTTCCTTCATTTTAAACAAACATTGAGACCTTA 537
Qy 238 TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn 257
Db 538 TACAATAACTTTAGTTAATAAATTCACGATTACTTAATTAATTAAGGCAAGATTAAAC 597
Qy 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277
Db 598 GATTGTAATGTTGAAAAAATGAACACATGTTAAATTAACATAACTTAGTGATTAAAA 657
Qy 278 AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaLysLysLys 297
Db 658 GCAATTGATGACAAATAGATCTTTTAAACCAATACGACTTCGAAGCAATTAATAAA 717
Qy 298 LeuIleAsnAspAspThrLysLysAspMetLeuGluLysLeuLeuSerThrGlyLeuVal 317
Db 718 TTGATAAATGATGATACGAAAAAAGATATGCTTGGCAAAATTAATTTAGTACAGGATTAGT 777
Qy 318 GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGluAspMetLeu 337
Db 778 CAAATTTTCTTAATACAAATATATCAAAATTAATTGAAGGAAAAATTCACAGATATGTTA 837
Qy 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357
Db 838 AACATTTCCAAACACCAATGCGTAAAAAACAATGTCAGAAAAATCTCGGATGTTTCAGA 897
Qy 358 HisLeuAspGluAtrGluGluCysLysCysLeuLeuAsnTyrLysGlnGluLysAspLys 377
Db 898 CATTTAGATGAAGAAGAAGATGTAATGTTTATTAATTAACAAACAAGAGGTGATAAA 957
Qy 378 CysValGluAsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAla 397
Db 958 TGTGTTGAAATCCAAATCCTTACTTGTAAACGAAATAATATGTTGGATGTGATCGACATGCC 1017
Qy 398 ThrCysThrGluGluAspSerGlySerSerArgLysLysIleThrCysGluCysThrLys 417
Db 1018 AAATGTACCCGAAGAAGATTTCAGGTAGCAACGGAAGAAAAATCACATGTGAATGTAATAA 1077
Qy 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSer 430
Db 1078 CCTGATCTTATCCACTTTTCGATGGTATTTCTGCGACT 1116
RESULT 39
AF040569 Plasmodium falciparum strain Kenya-3 merozoite surface protein-1
LOCUS (MSP-1) gene, partial cds.
DEFINITION
ACCESSION AF040569
VERSION AF040569.1 GI:2789669
KEYWORDS
SOURCE Plasmodium falciparum.
ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 1116)
AUTHORS Qari,S.H., Shi,Y.P., Goldman,I.F., Nahlen,B.L., Tibayrenc,M. and Lal,A.A.
TITLE Predicted and observed alleles of Plasmodium falciparum merozoite surface protein-1 (MSP-1), a potential malaria vaccine antigen
JOURNAL Mol. Biochem. Parasitol. 92 (2), 241-252 (1998)
MEDLINE 98319411
PUBMED 9657329
REFERENCE 2 (bases 1 to 1116)
AUTHORS Qari,S.H., Shi,Y.P., Goldman,I., Nahlen,B., Tibayrenc,M. and Lal,A.A.
TITLE Direct Submission
JOURNAL Submitted (31-DEC-1997) DASTLR, Centers for Disease Control & Prevention (CDC), 1600 Clifton Road, Atlanta, GA 30333, USA
FEATURES
source Location/Qualifiers
1..1116
/organism="Plasmodium falciparum"

```

```

/strain="Kenya-3"
/isolate="10"
/db_xref="taxon:5833"
/chromosome="9"
<1..>1116
/gene="MSP-1"
/allele="Kenya-3"
<1..>1116
/gene="MSP-1"
/feature="42 kDa C-terminal region"
/codon_start=1
/product="merozoite surface protein-1"
/protein_id="AAC39099.1"
/db_xref="GI:2789670"
/translation="AISVTMDNLSGFENEYDVIYKPLAGVYRSKLRQIKENFTFN
LNLDNLRLKRYFLVDLQKHISSNEYIIEEDSKLLNSKQNTLLKSYK
YKESVENDIKFAQEGISYEKVLAKYKDDLESKKVKEKEPSPPTPPSPAKT
DEQKESKELFELNIEYLNLYNKIDHYLINLAKINDCNVEKDEAHVITKLSDL
KALDDIKDLFRHNDFAIKKLLINDDKKMLKLLSTGLVONFPNTIISKLEGFQ
DMLNISQHCYKQCPENSGCFRHLDERECKLLNYKQEGDKCVENPNPTCNENNGG
CDAAKCYFEEDSGSGKKITCECTKPDSPYFPDFGIFCS"

```

BASE COUNT 485 a 145 c 160 g 326 t

ORIGIN

#### Alignment Scores:

Pred. No.: 1.07e-124 Length: 1116  
 Score: 1900.50 Matches: 365  
 Percent Similarity: 98.12% Conservative: 1  
 Best Local Similarity: 97.86% Mismatches: 6  
 Query Match: 83.54% Indels: 1  
 DB: 3 Gaps: 1

US-10-057-531A-2 (1-431) x AF040569 (1-1116)

```

Qy 58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77
Db 1 GCATATCTGTCAATGGATATATCTCTCAGGATTTTGAATAAGTAATGATGTTATA 60

Qy 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97
Db 61 TATTTAAACCTTTAGCTGGATATATACAGCTTAAAAAACAATTCGAAACAAACATT 120

Qy 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117
Db 121 TTTACATTTAAATTTAAATTTGAACGATATCTTAAATTCACGCTTAAAGAAACGAAATAT 180

Qy 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137
Db 181 TTCCTAGATCTATTAGAACTCTGATTTAATGCAATTTAAACATATATCTCTCAATGAATAC 240

Qy 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157
Db 241 ATTATTGAAGATTCATTTAAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTA 300

Qy 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluLysIleSer 177
Db 301 TACAAATATATAAAGAATTCAGTAGAAATGATATTAATTTTGCACAGGAAGGTATAAGT 360

Qy 178 TyrTyrGluLysValLeuAlaLysTyrLysAspLeuGluSerIleLysLysValIle 197
Db 361 TATTATCAAAAGGTTTACGGAATATAAGGATGATTAGATCAATTTAAAGAAAGTTATC 420

Qy 198 LysGluGluLysLysPheProSerSerProThrThrProProSerProAlaLys 217
Db 421 AAAGAAGAAAGAG---TTCCCATCATCACCAACAACACCTCCGTCACCAGCAAA 477

Qy 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237
Db 478 ACAGACGAACAAGAAAGAAAGTAAAGTTCCCTTCCATTTTAAACAACATTCAGACCTTA 537

Qy 238 TyrAspAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn 257
Db 538 TACAATAACTTGTATTAATAAATTTGACCATTTACTTAATTAAGTAAAGCAAGATTAAC 597

```

```

Qy 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277
Db 598 GATTGTGATGTTCAAAAGACATGAAGCACATGTTAAATAACTTAACCTAGTATTTAAA 657

Qy 278 AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaLysLys 297
Db 658 GCAATTTGATCACAATAATAGATCTTTTAAAAACCAATACGACTTCGAAGCAATTTAAA 717

Qy 298 LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal 317
Db 718 TTGATAAATGATGATACGAAAGAAAGATATGCTTGGCAAAATTTACTTAGTACAGGATGTT 777

Qy 318 GlnAsnPheProAsnThrIleIleSerLysLeuIleGluLysPheGlnAspMetLeu 337
Db 778 CAAAATTTTCTTAATAACAATAATATCAAAATTAATTTGAAGGAAATTTCCAAGATATGTTA 837

Qy 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357
Db 838 AACATTTTCAACACCAATCGCTTAAAAAACCAATGTCCAGAAATTTCTGGATGTTTTCAGA 897

Qy 358 HisLeuAspGluArgGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377
Db 898 CATTTAGATGAAAGAGAGAAATGTAATGTTTATTAATTTACAAACAAGAGGTGATAA 957

Qy 378 CysValGluAsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAla 397
Db 958 TGTGTTGAAATCCCAATCTACTTCTAAGCAAAATTAATGGTGTGATGCAGATGCC 1017

Qy 398 ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluGlyThrLys 417
Db 1018 AAATGTACCGAAGAGATTCAGGTAGCAACGGAAGAAATTCACATGCTGAATGTACTAAA 1077

Qy 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSer 430
Db 1078 CCTGATTCATTCATTTTCGATGTTATTTCTGTCAGT 1116

```

#### RESULT 40

PFAMSP807 1726 bp DNA linear INV 29-MAY-2002  
 P. falciparum DNA for the precursor to the major merozoite surface proteins, C-terminal.

DI3344  
 DI3344.1 GI:391786  
 EGF-like domains; major merozoite surface protein precursor.  
 Plasmodium falciparum (isolate 807 and 836) merozoite, DNA.  
 ORGANISM  
 Plasmodium falciparum  
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

REFERENCE  
 1 (bases 1 to 1726)  
 Jongwutives, S., Tanabe, K. and Kanbara, H.  
 Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP1) of Plasmodium falciparum from field isolates

Mol. Biochem. Parasitol. 59 (1), 95-100 (1993)

93295445

2 (bases 1 to 1726)

Jongwutives, S.

Direct Submission

Submitted (05-OCT-1992) Somchai Jongwutives, Institute of Tropical Medicine, Nagasaki Univ., Dept. of Protozoology, 12-4 Sakamoto-machi, Nagasaki 852, Japan (Tel:0958-47-2111(ex.3747),

Fax:0958-47-6607)

Submitted (05-OCT-1992) to DDBJ by:

Somchai Jongwutives  
 Department of Protozoology  
 Institute of Tropical Medicine  
 Nagasaki University  
 12-4 Sakamoto-machi  
 Nagasaki 852  
 Japan

Phone: 0958-47-2111x3747

Fax: 0958-47-6607.

Location/Qualifiers

#### FEATURES

```
source      1. .1726
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/chromosome="9"
/dev_stage="merozoite"
gene        1. .1712
<1. .1712
CDS         /gene="MSPI"
           /gene="MSPI"
           /codon_start=3
           /product="major merozoite surface protein precursor"
           /protein_id="BA02605.1"
           /db_xref="GI:391787"
           /translation="PLKTISEVSIQTDNYANLEKFRVLSKIDGKLNHLGKKKLS
           FSSGHLHILTELKYNKNTGNSPNNKKNVNEALKASYENFLPEAKVTVVTPPO
           ESDNDEYLDVVGTGSAISVTMDNIIISGRENEVDVLYLAPLAGVYSLKKQLEKNII
           FNLIANDLILNRLLKRRKYFLDVLESLDMOFKHISSENEYIIEISFKLNSEKQNLTKS
           YTKIKESVENDKFLPAGEGISYEKVLAKYKDDLESIKVKIEKEPEPSPPTPPSPA
           KTDQKESKFLPFLNIETLYNNLVNKTGDYLIINLAKINDCNVKEDEAHVKITKLS
           DLKAIIDKIDLFKNTDDEFAIKKLIINDTKDKMLKLLSTGLVQNPPTIISKLEIGK
           FQMLNISQHCQVKOCPONSGCFRHLDERECKCLLVKQEGDKGVENPNPTCNENN
           GGCDADAKTEEDSGSNGKKTCECTKPDSPYFPFDGIFCSSSNFLGISFLILMLILY
           SFI"
BASE COUNT 739 a 233 c 246 g 508 t
ORIGIN
Alignment Scores:
Pred. No.: 1-82e-124 Length: 1726
Score: 1900.00 Matches: 367
Percent Similarity: 95.43% Conservative: 9
Best local Similarity: 93.15% Mismatches: 12
Query Match: 83.52% Indels: 6
DB: 3 Gaps: 2

US-10-057-531A-2 (1-431) x PFAMSP807 (1-1726)
QY 38 ThrAspAspAspLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMet 57
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 486 TCCGAGATAATGACGAATATTAGATCAAGTAGTACTGGAGAA----- 530
QY 58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 531 GCATATCTGTGCAATGGAATAATATCTCTCAGGATTTTGAATAATGAATATGATGTTATA 590
QY 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 591 TATTTAAACCTTTAGCTGGAGTATATAGAGCTTAAAAAACAAATTTGAAAAACATT 650
QY 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 651 ATTACATTTAAATTTGAACGATATCTTAAATTCAGCTCTTAAGAAACGAAAAATAT 710
QY 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 711 TTCTTAGATGTTAGTAATCTGATTTAATGCAATTTAAACATATATCTCCTCAATGAATAC 770
QY 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 771 ATTATTCAAGATTCAATTTAAATTTAATTCAGAAACAAAAAACACACACATTTTAAAAAGT 830
QY 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 831 TACAAATATATAAAGAAATCAGTAGAAATGATATTAAATTTGCACAGGAGGTATAAGT 890
QY 178 TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysValIle 197
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 891 TATTATGAAAAGGTTTTCAGCAATATATAGCATGATTTAGATCAATTTAAAAAGTTATC 950
QY 198 LysGluGluLysGluLysPheProSerSerProProThrThrProProSerProAlaLys 217
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 951 AAGAAGAAAAAGGAG---TTCCCATCATCATCACCACCAACACACCTCCGTCACCAGCAAAA 1007
```

```
QY 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1008 ACAGACCAACAAAAAGAGAAAGTAAGTTCCTTCCATTTTAAACAAACATTGAGACCTTA 1067
QY 238 TyrAsnAsnLeuValAsnLysIleAspAspTyrIleLeuIleAsnLeuLysAlaLysIleAsn 257
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1068 TACAATAAATCTAGTTAATAAAATTTGGCGATTACTTAATTAACTTTAAAGGCAAGATTAAAC 1127
QY 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1128 GATTGTAAATGTTGAAAAGAGTAGACACATGTTAAATACTAACTTAGTGATTTAAAA 1187
QY 278 AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys 297
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1188 GCAATTCATGACAAAATAGATCTTTTTTAAAAACACTCACGACTTCGAAGCAATTTAAAAA 1247
QY 298 LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal 317
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1248 TTGATAAATCATGATACGAAAAAGATATGCTTGGCAAAATTACTTAGTACAGGATTAGTT 1307
QY 318 GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeu 337
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1308 CAAATTTTCCCTAATACATAATATCAAAATTAATTCGAAGAAAATTTCCAGATATGTTA 1367
QY 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1368 AAGATTTTCAACACCAATCGTAAAAAAACAAATGTCCACAAAATTTCTGGATGTTTCAGA 1427
QY 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1428 CATTTAGATGAAGAGAGAAATGTAATGTTTAAATTTACAAACAAAGAGGTGATATAA 1487
QY 378 CysValGluAsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAla 397
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1488 TGTGTTGAAAATCCAAATCCTACTTGTACGAAAATAATGTTGATGTCAGATGCC 1547
QY 398 ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1548 AAATGTACCCGAAGAAGATTTCAGGTAGCAGGAAAGAAAATTCACATGTGAATGTACTAAA 1607
QY 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1608 CCGATTTCTTATCCATTTTTCGATGGGTATTTTCTGCGAGTTCC 1649
```

Search completed: May 19, 2003, 15:54:17

Job time : 2641 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

# OM protein - protein search, using sw model

Run On: May 12, 2003, 10:21:03 ; Search time 39 Seconds  
(without alignments)  
1472.590 Million cell updates/sec

Title: US-10-057-531A-2

Perfect score: 431

Sequence: 1 MHHHHHHHSLVPRGSMKE.....TCECTKPDVPLFDGIFCSS 431

Scoring table:

OLIGO Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : A\_Geneseq\_101002.\*

```
1: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	375	87.0	375	22	AA83926
2	60	13.9	116	18	AAW36103
3	60	13.9	116	18	AAW22592
4	60	13.9	127	18	AAW22593
5	60	13.9	127	18	AAW36102
6	51	11.8	54	14	AAW41357
7	50	11.6	76	21	AAW26110
8	50	11.6	167	22	AAW85253
9	50	11.6	167	22	AAW85254
10	50	11.6	289	22	AAE13033

11	50	11.6	289	23	AAW47732	HOP38 #5. Helicob
12	50	11.6	402	23	ABG67266	Polythoa 2 fluores
13	50	11.6	407	23	ABG67267	Polythoa 2 fluores
14	50	11.6	407	23	ABG67270	Polythoa 2 fluores
15	50	11.6	467	21	AAW85150	Mouse secreted pro
16	50	11.6	467	21	AAW79067	Murine secreted pr
17	50	11.6	551	22	AAW85251	Plant thioredoxin-
18	50	11.6	551	22	AAW85252	Plant thioredoxin-
19	50	11.6	692	21	AAW11532	SEN virus protein
20	50	11.6	824	23	AAW48247	Thioredoxin/deoxyx
21	48	11.1	49	14	AAW41355	MSPIEGF1B EGPI-lik
22	46	10.7	96	22	AAW37608	Merzoite surface
23	46	10.7	108	22	AAW37609	Merzoite surface
24	46	10.7	138	21	AAW11534	SEN virus protein
25	46	10.7	284	23	AAE20460	Recombinant (SPKR)
26	46	10.7	355	20	AAW09372	Merzoite surface
27	46	10.7	355	20	AAW05832	Merzoite surface
28	46	10.7	361	20	AAW09373	Merzoite surface
29	46	10.7	361	20	AAW05833	Merzoite surface
30	46	10.7	376	20	AAW09374	Modified merzoite
31	46	10.7	376	20	AAW05834	Modified merzoite
32	46	10.7	408	20	AAW84315	TrxA-rabbit tissue
33	46	10.7	1639	19	AAW54145	P. faiciparum synt
34	45	10.4	59	23	ABW80661	N-His botulinum to
35	45	10.4	218	21	AAW11533	SEN virus protein
36	34	7.9	49	14	AAW41354	MSPIEGF1A EGPI-lik
37	32	7.4	55	22	AAW51128	Human mammaglobin
38	27	6.3	382	20	AAW39952	Gaussia luciferase
39	27	6.3	382	23	AAE13383	Gaussia species CB
40	23	5.3	54	14	AAW41356	MSPIEGF2A EGPI-lik
41	23	5.3	1654	6	AAW50777	Sequence of the p1
42	21	4.9	299	21	AAW32143	Alfa transcript
43	21	4.9	302	21	AAW08323	Amino acid sequenc
44	21	4.9	328	22	AAE05749	Chimeric S peptide
45	18	4.2	18	21	AAW70287	Plasmodium falcipa
46	18	4.2	350	21	AAW70278	Recombinant vaccin
47	17	3.9	17	21	AAW70289	Plasmodium falcipa
48	17	3.9	24	23	AAW7402	Amino acid sequenc
49	17	3.9	131	10	AAW90410	Plasmodium falcipa
50	16	3.7	31	19	AAW1706	S-peptide with (g1
51	15	3.5	15	18	AAW08161	Purification tag o
52	15	3.5	15	19	AAW79592	BS15-peptide. Syn
53	15	3.5	15	19	AAW79592	Peptide comprising
54	15	3.5	15	20	AAW84204	S-tag sequence. U
55	15	3.5	15	21	AAW99879	Bovine pancreatic
56	15	3.5	15	23	AAE20474	Rec2 protein leade
57	15	3.5	19	17	AAW00339	S-peptide, a known
58	15	3.5	20	18	AAW42243	Leader sequence fo
59	15	3.5	20	18	AAW08594	Ribonuclease S-pep
60	15	3.5	20	20	AAW04117	Ribonuclease S-pep
61	15	3.5	20	21	AAW99679	N-terminal His-tag
62	15	3.5	20	21	AAW28485	Human and murine O
63	15	3.5	20	21	AAW87744	Recombinant lysoph
64	15	3.5	20	22	AAW82670	Integrin cytoplasm
65	15	3.5	20	22	AAW50878	N-terminal sequenc
66	15	3.5	20	22	AAW59137	His tag peptide SE
67	15	3.5	20	23	ABW94460	Hexa-His/thrombin
68	15	3.5	21	16	AAW87745	Leader sequence fo
69	15	3.5	21	18	AAW08590	Fusion peptide lea
70	15	3.5	21	20	AAW07403	Fusion peptide lea
71	15	3.5	21	20	AAW97132	OB gene-related pe
72	15	3.5	21	21	AAW28465	OB gene-related pe
73	15	3.5	21	21	AAW28484	N-terminal extensi
74	15	3.5	21	21	AAW26857	Human and murine O
75	15	3.5	21	23	AAW75380	Immunotoxin fusion
76	15	3.5	22	18	AAW35362	Synthetic 65k:glut
77	15	3.5	23	19	AAW65551	Peptide expressed
78	15	3.5	23	21	AAW85288	Peptide used in th
79	15	3.5	25	17	AAW92906	His-Tag leader seq
80	15	3.5	25	20	AAW32089	Toscana virus reco
81	15	3.5	28	22	AAE04756	Toxoplasma gondii
82	15	3.5	43	17	AAW92733	PET-15b expression
83	15	3.5	43	21	AAW28452	Murine OB carboxy

84	15	3.5	43	21	AAB28471	Murine OB peptide	157	15	3.5	371	20	AAW73369	Epitope tagged TBP
85	15	3.5	43	21	AAW77930	Murine OB protein	158	15	3.5	386	23	AAE22205	His6-NLS-Cre-MTS f
86	15	3.5	110	19	AAW78908	Rat CART1 fusion p	159	15	3.5	398	19	AAW71071	Multiple sclerosits
87	15	3.5	110	20	AAV45187	pET15b-rCART1 fusi	160	15	3.5	398	20	AAW99557	Protein encoded by
88	15	3.5	116	21	AAV52367	Human tyrosine kin	161	15	3.5	398	23	ABW84070	Human proBMP-2. H
89	15	3.5	121	22	AAB31608	Amino acid sequenc	162	15	3.5	409	21	AAW07702	Protein encoded by
90	15	3.5	124	16	AAW49151	Wild-type bovine p	163	15	3.5	409	22	AAE02201	Amino acid sequenc
91	15	3.5	124	18	AAW35108	Bovine pancreatic	164	15	3.5	431	22	AAE02201	Streptococcus pneu
92	15	3.5	124	20	AAV33327	Bovine pancreatic	165	15	3.5	431	22	AAE02201	Streptococcus pneu
93	15	3.5	124	20	AAW81768	Bovine RNAS peptid	166	15	3.5	485	22	AAU00222	LFN-Bcl-XL apoptos
94	15	3.5	124	21	AAV69974	RNAS protein. Bos	167	15	3.5	588	21	AAW01279	Histidine tagged C
95	15	3.5	133	20	AAV33443	Parvovirus rppv	168	15	3.5	570	21	AAW01276	Histidine tagged A
96	15	3.5	144	19	AAW47195	Herpes simplex vir	169	15	3.5	656	23	AAU75383	Immunotoxin fusion
97	15	3.5	144	21	AAV52368	Human tyrosine kin	170	15	3.5	657	23	AAU75389	Diphtheria toxin/U
98	15	3.5	149	23	AAU10045	Tm 2.2 clone/His t	171	15	3.5	662	21	AAW01273	Neisseria meningit
99	15	3.5	149	23	AAU10047	Tm 2.3 clone/His t	172	15	3.5	665	20	AAV24940	HCV NS4A-NS3 compl
100	15	3.5	149	23	AAU10049	Tm 13.17 clone/His	173	15	3.5	665	20	AAV24941	HCV NS4A-NS3 compl
101	15	3.5	149	23	AAU10051	Tm 3.4 clone/His t	174	15	3.5	665	20	AAV24942	HCV NS4A-NS3 compl
102	15	3.5	149	23	AAU10053	Tm 3.9 clone/His t	175	15	3.5	665	20	AAV24943	HCV NS4A-NS3 compl
103	15	3.5	149	23	AAU10055	Tm 7.5 clone/His t	176	15	3.5	665	20	AAV24944	HCV NS4A-NS3 compl
104	15	3.5	150	21	AAV52369	Human tyrosine kin	177	15	3.5	665	20	AAV24945	HCV NS4A-NS3 compl
105	15	3.5	150	23	AAE20141	Human tyrosine kin	178	15	3.5	665	20	AAV24946	HCV NS4A-NS3 compl
106	15	3.5	152	23	AAW50302	Factor XIIa subst	179	15	3.5	665	20	AAV24947	HCV NS4A-NS3 compl
107	15	3.5	153	22	AAW50303	Amino acid sequenc	180	15	3.5	669	23	ABW08420	Catalytic trans-s
108	15	3.5	156	23	AAW50303	Nerve growth facto	181	15	3.5	671	20	AAV24948	HCV NS4A-NS3 compl
109	15	3.5	157	21	AAW01596	Nerve growth facto	182	15	3.5	671	20	AAV24949	HCV NS4A-NS3 compl
110	15	3.5	157	22	AAW67677	Amino acid sequenc	183	15	3.5	711	19	AAW57321	Topoisomerase I ex
111	15	3.5	157	22	AAW67677	Synthetic nerve gr	184	15	3.5	948	22	AAW31611	Amino acid sequenc
112	15	3.5	160	22	AAW50300	Streptococcus pneu	185	14	3.2	18	20	AAW73374	Epitope tagged TBP
113	15	3.5	161	23	ABW81281	Human FGF-9 mutant	186	13	3.0	15	21	AAW06025	Bovine truncated S
114	15	3.5	162	23	AAW50300	Factor XIIa subst	187	13	3.0	18	10	AAW91370	Apomigrin peptide
115	15	3.5	163	22	AAW67678	Amino acid sequenc	188	12	2.8	12	20	AAW94422	MBP/BMP fusion con
116	15	3.5	166	17	AAW00525	Human obesity prot	189	12	2.8	18	13	AAW25208	MBP/BMP fusion con
117	15	3.5	166	17	AAW00529	Murine obesity prot	190	12	2.8	18	13	AAW25208	Bovine seminal RNA
118	15	3.5	166	17	AAW00533	Murine obesity prot	191	12	2.8	18	21	AAV70288	hnRNP U protein 69
119	15	3.5	166	17	AAW00534	Murine obesity prot	192	12	2.8	19	21	AAV70288	EGF-like domain of
120	15	3.5	166	17	AAW00534	Murine obesity prot	193	12	2.8	20	22	AAW70818	Extension peptide
121	15	3.5	166	22	AAW02196	Staphylococcus aur	194	12	2.8	21	20	AAW18411	Polyptide which
122	15	3.5	166	23	AAW50301	Nerve growth facto	195	12	2.8	24	22	AAW70817	Sequence #2 for pa
123	15	3.5	167	17	AAW92730	Murine obesity prot	196	12	2.8	24	22	AAW70817	Aminoacyl tRNA syn
124	15	3.5	167	17	AAW00521	Human obesity prot	197	12	2.8	134	20	AAW73326	Modified polyHis-e
125	15	3.5	167	17	AAW92726	Human obesity prot	198	11	2.6	11	22	AAW83927	Recombinant squir
126	15	3.5	167	22	AAW00517	Human obesity prot	199	10	2.3	19	16	AAW77413	Recombinant squir
127	15	3.5	167	22	AAW67679	Amino acid sequenc	200	9	2.1	20	7	AAW60976	Recombinant squir
128	15	3.5	173	23	AAU10044	Tm 2.2 clone-His t	201	9	2.1	14	23	AAU77404	Amino acid sequenc
129	15	3.5	173	23	AAU10046	Tm 3.4 clone/His t	202	9	2.1	16	23	AAW98499	His-tag peptide SE
130	15	3.5	173	23	AAU10050	Tm 3.9 clone/His t	203	9	2.1	21	18	AAW23647	Polyhistidine-ente
131	15	3.5	173	23	AAU10052	Tm 7.5 clone/His t	204	9	2.1	21	18	AAW23647	Polyhistidine-ente
132	15	3.5	174	23	AAU10054	Tm 13.17 clone/His	205	9	2.1	21	18	AAW23650	Met Gly His10 pept
133	15	3.5	180	19	AAW47081	Salmonella Sef14 p	206	9	2.1	22	22	AAW23653	12 amino acid ente
134	15	3.5	181	23	AAU10048	Tm 13.17 clone/His	207	9	2.1	22	22	AAW49518	Enterokinase cleav
135	15	3.5	189	22	AAE09217	Hexa-His-tagged hu	208	9	2.1	24	18	AAW19779	Peptide for expres
136	15	3.5	206	21	AAW44729	Hepatitis C virus	209	9	2.1	24	20	AAW06647	ARH-III angiogenin
137	15	3.5	213	20	AAV17898	HCV NS4A-NS3 compl	210	9	2.1	24	20	AAV27015	HFRT peptide. Syn
138	15	3.5	213	20	AAV17887	HCV NS4A-NS3 compl	211	9	2.1	24	20	AAV27015	HFRT peptide. Syn
139	15	3.5	215	20	AAV17890	HCV NS4A-NS3 compl	212	9	2.1	24	21	AAW81959	Recombinat antic
140	15	3.5	216	20	AAV17879	HCV NS4A-NS3 compl	213	9	2.1	24	22	AAW47057	Recombinat neutro
141	15	3.5	216	20	AAV17880	HCV NS4A-NS3 compl	214	9	2.1	26	12	AAW13621	S. epidermidis ope
142	15	3.5	216	20	AAV17881	HCV NS4A-NS3 compl	215	9	2.1	26	12	AAW13621	Staphylococcus epi
143	15	3.5	216	20	AAV17882	HCV NS4A-NS3 compl	216	9	2.1	53	19	AAW36986	Amino acid sequenc
144	15	3.5	216	20	AAV17883	HCV NS4A-NS3 compl	217	9	2.1	54	19	AAW36990	Amino acid sequenc
145	15	3.5	216	20	AAV17884	HCV NS4A-NS3 compl	218	9	2.1	91	20	AAW09127	Histidine tagged-h
146	15	3.5	216	20	AAV17885	HCV NS4A-NS3 compl	219	9	2.1	93	20	AAW09128	His-tagged Sonic h
147	15	3.5	216	20	AAV17886	HCV NS4A-NS3 compl	220	9	2.1	134	22	AAW82144	HIV TAT transduct
148	15	3.5	216	20	AAV17887	HCV NS4A-NS3 compl	221	9	2.1	139	23	ABP39568	B. steierthermophi
149	15	3.5	216	20	AAV17878	HCV NS4A-NS3 compl	222	9	2.1	183	21	AAW84920	Protein sequence o
150	15	3.5	217	21	AAV44730	Hepatitis C virus	223	9	2.1	183	21	AAW70866	B. mori 10XHis-TRA
151	15	3.5	220	21	AAW86630	Cell factor deriva	224	9	2.1	192	21	AAW92865	P. mori TRAS1 EN p
152	15	3.5	239	21	AAW52366	Human tyrosine kin	225	9	2.1	192	21	AAW92865	PF505 protein cont
153	15	3.5	248	23	ABW06951	His-OST311 protein	226	9	2.1	195	21	AAW11377	
154	15	3.5	262	22	AAW37610	Mercozite surfacis	227	9	2.1	254	18	AAW37699	
155	15	3.5	288	23	AAW50707	Bacillus anthracis	228	9	2.1	269	23	AAW49520	
156	15	3.5	333	22	AAW02193	Staphylococcus aur	229	9	2.1	323	19	AAW71185	

230	9	2.1	332	22	AAE02191	Staphylococcus aur	303	7	1.6	10	22	AAB75181	Recombinant A33 an
231	9	2.1	364	21	AAG13011	Q-tagged FabH prot	304	7	1.6	10	22	AAB60389	Thrombin substrate
232	9	2.1	393	22	AAG68149	Streptomyces grise	305	7	1.6	11	20	AAW93048	Peptide 1f-trb-eps
233	9	2.1	393	22	AAB74844	S. griseus actinom	306	7	1.6	11	20	AAW93031	Peptide 1f-trb-eps
234	9	2.1	411	22	AAU00219	Bcl-X1-DK apoptos	307	7	1.6	11	21	AAW11895	H3SA mutant His-ta
235	9	2.1	423	19	AAW68098	Chlamydomonas rein	308	7	1.6	12	21	AAW53265	pIII coat protein
236	9	2.1	448	19	AAW68399	Clostridium botuli	309	7	1.6	12	22	AAU04934	Human plasminogen
237	9	2.1	451	19	AAW68395	Clostridium botuli	310	7	1.6	12	22	AAU04935	Human plasminogen
238	9	2.1	451	19	AAW68398	Clostridium botuli	311	7	1.6	12	22	AAU04935	Human plasminogen
239	9	2.1	452	19	AAW68396	Clostridium botuli	312	7	1.6	13	23	AAU77403	Sequence #1 for pa
240	9	2.1	462	17	AAW95009	Type A neurotoxin	313	7	1.6	13	23	AAU09365	N-terminal His6 ta
241	9	2.1	462	19	AAW68390	Clostridium botuli	314	7	1.6	14	21	AAW23044	Endostatin vector
242	9	2.1	462	19	AAW68397	Clostridium botuli	315	7	1.6	14	21	AAW11893	Cyclo-[Cys-His-Met
243	9	2.1	472	19	AAW68394	Clostridium botuli	316	7	1.6	14	21	AAW66810	T cell antigen rec
244	9	2.1	472	19	AAW68393	Clostridium botuli	317	7	1.6	15	22	AAW68977	pGEX-7Hpro proteas
245	9	2.1	473	19	AAW68400	Clostridium botuli	318	7	1.6	17	20	AAW95000	N-terminal histidi
246	9	2.1	491	23	AAW68400	Clostridium botuli	319	7	1.6	17	23	AAO21607	Murine Toll-like r
247	9	2.1	567	22	AAU00220	Herbicidally activ	320	7	1.6	19	21	AAW90748	Protease Kex-2 cle
248	9	2.1	600	23	AAE20482	Bad-DTR apoptosis	321	7	1.6	19	21	AAW84796	Polypeptide fragme
249	8	1.9	8	21	AAE19475	Human protein for	322	7	1.6	21	7	AAW60028	Sequence of identi
250	8	1.9	11	22	AAW83928	Peptide derived fr	323	7	1.6	21	22	AAO5538	Human polypeptide
251	8	1.9	14	23	ABG67084	EGF-like domain of	324	7	1.6	23	23	AAO19984	Peptide of a prote
252	8	1.9	17	20	AAW03939	Streptavidin-bindi	325	7	1.6	24	22	AAW64594	OmpC-(6His)1. Syn
253	8	1.9	17	20	ABW08913	T2 peptide, synthe	326	7	1.6	24	22	AAO55865	Human polypeptide
254	8	1.9	23	18	AAW5248	TNF/hPI fusion pla	327	7	1.6	29	20	AAW43092	Tag sequence used
255	8	1.9	31	21	AAW15670	N-terminal peptide	328	7	1.6	32	23	ABW84805	Escherichia coli a
256	8	1.9	60	17	AAW90286	Alpha V beta 3 rec	329	7	1.6	37	22	AAW64595	Betalaphal/MBP-55
257	8	1.9	60	18	AAW20250	Nickel binding pro	330	7	1.6	37	22	AAW64595	OmpC-(6His)2. Syn
258	8	1.9	61	21	AAW24617	H. pylori cytoplas	331	7	1.6	41	21	AAW26109	H. contortus clone
259	8	1.9	61	21	AAW14344	H. pylori cytoplas	332	7	1.6	42	23	AAO19986	Protein of (double
260	8	1.9	73	23	ABW80911	Bovine pancreatic	333	7	1.6	42	23	AAO19986	Protein of (double
261	8	1.9	78	18	AAW20756	Murine eotaxin mat	334	7	1.6	45	14	AAW42283	Corticosterone. Sy
262	8	1.9	97	23	ABW80897	H. pylori cytoplas	335	7	1.6	50	22	AAW47506	Propionibacterium
263	8	1.9	98	23	ABW80898	Murine eotaxin pol	336	7	1.6	54	23	ABP31575	OmpC-(6His)3. Syn
264	8	1.9	123	18	AAW28331	Murine eotaxin pol	337	7	1.6	54	23	ABP31575	Human kinase-like
265	8	1.9	131	22	AAW81934	Staphylococcus aur	338	7	1.6	62	23	AAW85598	Ral2S-L985PEX pept
266	8	1.9	149	21	AAW08634	S. epidermidis ope	339	7	1.6	64	16	AAW82537	Hybrid IA beta cha
267	8	1.9	155	21	AAW08633	Arabidopsis thalia	340	7	1.6	65	16	AAW82535	Hybrid IE beta cha
268	8	1.9	155	21	AAW43258	Arabidopsis thalia	341	7	1.6	65	20	AAW94263	H6FXTN12 fusion pr
269	8	1.9	158	21	AAW08632	Arabidopsis thalia	342	7	1.6	68	22	AAW91324	Human immune/haema
270	8	1.9	158	21	AAW43257	Arabidopsis thalia	343	7	1.6	68	23	ABP02234	Human OREX protein
271	8	1.9	159	21	AAW43256	Arabidopsis thalia	344	7	1.6	77	22	AAW55128	Propionibacterium
272	8	1.9	244	21	AAW93977	Synthetic green fl	345	7	1.6	79	21	AAW02158	Human secreted pro
273	8	1.9	244	21	AAW93978	Synthetic green fl	346	7	1.6	80	22	AAU64509	Propionibacterium
274	8	1.9	342	23	ABG33381	Human Trx_Wt1_B.	347	7	1.6	82	22	ABW27363	Human peptide #14
275	8	1.9	379	23	ABG33382	Lactococcus lactis	348	7	1.6	82	22	ABW32509	Peptide #15 encode
276	8	1.9	391	18	AAW23534	CDK inhibitory fus	349	7	1.6	82	22	ABW18015	Peptide #14 encode
277	8	1.9	391	20	AAW95094	Human p27-p16 fusi	350	7	1.6	82	22	AAW53343	Human brain expres
278	8	1.9	391	21	AAW97526	Human W3 protein s	351	7	1.6	82	22	AAW53343	Human bone marrow
279	8	1.9	391	21	AAW96041	Antiproliferative	352	7	1.6	82	22	AAW13580	Peptide #14 encode
280	8	1.9	391	21	AAW96068	Angiogenesis inhib	353	7	1.6	82	22	AAW25978	Peptide #15 encode
281	8	1.9	410	23	ABG33382	Human Trx_Wt1_A.	354	7	1.6	82	22	AAW01333	Peptide #15 encode
282	8	1.9	416	21	AAW18185	Plasmodium falcipa	355	7	1.6	89	22	ABG33350	Human peptide enco
283	8	1.9	431	22	ABW69274	Drosophila melanog	356	7	1.6	91	23	ABW54331	OmpC-(6His)6. Syn
284	8	1.9	544	23	ABG40264	Human peptide enco	357	7	1.6	91	23	ABW54331	Lactococcus lactis
285	8	1.9	591	23	ABG33383	Human Trx_Wt1. Ho	358	7	1.6	94	21	AAW13471	Lactococcus lactis
286	8	1.9	622	21	AAW97170	Human FGF-RI Extra	359	7	1.6	94	22	AAW83147	Protein encoded by
287	8	1.9	763	21	AAW38942	Arabidopsis thalia	360	7	1.6	94	23	ABW94118	Protein encoded by
288	8	1.9	887	22	AAE07889	Modified clostridi	361	7	1.6	98	22	AAU42269	Chlamydia protein
289	8	1.9	887	22	AAE07892	Modified clostridi	362	7	1.6	98	22	AAU42269	Propionibacterium
290	8	1.9	974	22	ABW62110	Drosophila melanog	363	7	1.6	99	22	ABW33147	Peptide #6653 enco
291	8	1.9	1059	21	AAW93309	A manganese supero	364	7	1.6	99	22	AAW59798	Human brain expres
292	8	1.9	1067	21	AAW93307	A manganese supero	365	7	1.6	99	22	AAW72385	Human bone marrow
293	8	1.9	1070	21	AAW93308	A manganese supero	366	7	1.6	99	23	ABG42203	Peptide #6676 enco
294	8	1.9	1084	21	AAW93312	A manganese supero	367	7	1.6	99	23	ABG42203	Human peptide enco
295	8	1.9	1092	21	AAW93310	A manganese supero	368	7	1.6	112	23	ABG42203	Human Wt1-Tr4. Ho
296	8	1.9	1095	21	AAW93311	A manganese supero	369	7	1.6	117	20	AAW21867	Amino acid sequenc
297	7	1.6	7	9	AAW80401	Affinity peptide c	370	7	1.6	118	22	AAU31100	Novel human secret
298	7	1.6	7	23	AAE20476	Bovine ribonucleas	371	7	1.6	119	22	ABW35472	Peptide #2954 enco
299	7	1.6	8	15	AAW57097	N-terminal fragmen	372	7	1.6	119	22	ABW35472	Peptide #2978 enco
300	7	1.6	8	16	AAW77463	GST-I-domain fusio	373	7	1.6	119	22	AAW56291	Protein #2910 enco
301	7	1.6	8	16	AAW48167	Histidine tag to c	374	7	1.6	119	22	AAW56291	Human brain expres
302	7	1.6	10	22	AAW98321	Recombinant A33 an	375	7	1.6	119	22	AAW16487	Human bone marrow

376	7	1.6	119	22	AAW28977	Peptide #3014 enco	449	7	1.6	299	23	ABB75000	Ra12/C-terminal po
377	7	1.6	119	22	AAW04212	Peptide #2894 enco	450	7	1.6	299	23	AAE17587	Mycobacterium spec
378	7	1.6	119	23	ABG38250	Human peptide enco	451	7	1.6	302	21	AAV96926	M. tuberculosis an
379	7	1.6	120	20	AAV21857	Amino acid sequenc	452	7	1.6	304	22	AAU69902	Human /M. tubercul
380	7	1.6	121	20	AAV21858	Amino acid sequenc	453	7	1.6	304	22	AAW01257	Human /M. tubercul
381	7	1.6	128	18	AAW08427	Vector expressing	454	7	1.6	304	23	ABW95362	Ra12-P775P-ORF3 co
382	7	1.6	128	20	AAV30055	Amino acid sequenc	455	7	1.6	306	22	AAW67481	Ra12- P775P-ORF3 c
383	7	1.6	129	21	AAV31473	protein encoded by	456	7	1.6	306	22	AAU74596	A fusion protein c
384	7	1.6	129	22	ABG83149	Chlamydia protein	457	7	1.6	314	21	AAW67481	Antigenic fusion p
385	7	1.6	129	23	ABG94120	TGF-beta fusion pr	458	7	1.6	314	22	AAW33224	Eucahyptus grandis
386	7	1.6	130	18	AAW08185	Arabidopsis thalia	459	7	1.6	314	22	AAE11933	Human CG27 (or C86
387	7	1.6	130	21	AAW34414	heGF-ST-IGF-1 fusi	460	7	1.6	320	21	AAW33224	Lung tumour protei
388	7	1.6	131	15	AAW63193	Breast and ovarian	461	7	1.6	320	21	AAW33224	Recombinant gp 90
389	7	1.6	131	21	AAW58736	Protein encoded by	462	7	1.6	324	22	ABG20509	Novel human diagno
390	7	1.6	132	22	AAW83268	Chlamydia protein	463	7	1.6	327	22	ABG69341	C. trachomatis hyp
391	7	1.6	133	23	ABW94239	TGF-beta fusion pr	464	7	1.6	330	22	AAU38918	Antigenic fusion p
392	7	1.6	134	18	AAW08187	TGF-beta fusion pr	465	7	1.6	330	23	AAU74589	Mycobacterium sp.
393	7	1.6	136	18	AAW08186	Protein library re	466	7	1.6	330	23	AAE17566	Mycobacterium spec
394	7	1.6	136	23	ABW06192	Human polypeptide	467	7	1.6	330	23	AAE17567	Mycobacterium tube
395	7	1.6	137	22	AAW04232	M. tuberculosis an	468	7	1.6	331	20	AAV32060	Pseudomonas aerugi
396	7	1.6	137	22	AAW01885	Human breast cance	469	7	1.6	331	22	AAW66450	Mycobacterium tube
397	7	1.6	158	22	AAW63415	Gene 30 human secr	470	7	1.6	332	18	AAW32418	Mycobacterium tube
398	7	1.6	162	22	AAU01687	Histidine-tagged t	471	7	1.6	332	18	AAW32250	M. tuberculosis lm
399	7	1.6	164	22	AAW82901	OmpC-(GHis)12. Sy	472	7	1.6	332	19	AAW81683	M. tuberculosis an
400	7	1.6	167	22	AAW64598	Parathyroid hormon	473	7	1.6	332	19	AAW64322	M. tuberculosis re
401	7	1.6	170	23	AAU11950	Human polypeptide	474	7	1.6	332	20	AAV39083	M. tuberculosis re
402	7	1.6	171	22	AAO13893	Amino acid sequenc	475	7	1.6	332	20	AAV38945	M. tuberculosis op
403	7	1.6	184	22	AAW30581	Human WTL-B. Homo	476	7	1.6	332	22	AAU01895	Mycobacterium spec
404	7	1.6	188	23	ABG33385	Murine endostatin.	477	7	1.6	332	23	AAE17584	Arabidopsis thalia
405	7	1.6	191	21	AAW28398	Amino acid sequenc	478	7	1.6	333	21	AAW31543	Heat-resistant ger
406	7	1.6	191	23	AAW77950	S. carnosus nitrat	479	7	1.6	334	18	AAW22517	Helicobacter pylor
407	7	1.6	196	18	AAW26746	H6FTXN123 fusion p	480	7	1.6	344	22	AAW66427	Helicobacter pylor
408	7	1.6	197	20	AAW94262	Protein encoded by	481	7	1.6	344	22	AAW66448	Arabidopsis thalia
409	7	1.6	202	21	AAW13474	Chlamydia protein	482	7	1.6	348	21	AAW30654	Arabidopsis thalia
410	7	1.6	202	22	AAW83150	Leishmania antigen	483	7	1.6	348	22	ABW68968	Drosophila melanog
411	7	1.6	202	23	AAW94121	Leishmania antigen	484	7	1.6	351	20	AAV25407	Human tissue facto
412	7	1.6	206	19	AAW70222	Leishmania antigen	485	7	1.6	352	22	AAW66449	Pseudomonas aerugi
413	7	1.6	206	23	ABW60874	Leishmania antigen	486	7	1.6	358	22	AAU00214	Fusion protein GST
414	7	1.6	206	23	AAU71818	Leishmania antigen	487	7	1.6	360	21	AAW32575	Eucalyptus grandis
415	7	1.6	210	16	AAW85066	Wilms' tumour anti	488	7	1.6	363	19	AAW42427	PDC-E2, BCOADC-E2,
416	7	1.6	210	18	AAW22883	Wilms' tumour prot	489	7	1.6	364	20	AAW07329	Recombinant T vari
417	7	1.6	212	21	AAW30656	Arabidopsis thalia	490	7	1.6	368	20	AAV32069	Mycobacterium tube
418	7	1.6	219	22	AAU01884	M. tuberculosis an	491	7	1.6	368	23	AAU74598	Antigenic fusion p
419	7	1.6	220	22	AAO22141	Ra12-mammaglobin f	492	7	1.6	369	23	ABG33390	Human WTL-C. Homo
420	7	1.6	224	18	AAW14788	FKBP-LCK:SH2 fusio	493	7	1.6	374	22	ABW58849	Drosophila melanog
421	7	1.6	224	20	AAW96823	A fusion protein o	494	7	1.6	378	18	AAW14786	FKBP-ZAP:SH2 fusio
422	7	1.6	224	22	AAU69899	Human prostate pro	495	7	1.6	378	23	ABW6821	A fusion protein o
423	7	1.6	224	22	AAU01254	Ra12-P510S-C const	496	7	1.6	378	23	ABW77877	Sequence of trunca
424	7	1.6	224	23	ABW95359	Ra12- P510S-C const	497	7	1.6	384	18	AAW14787	FKBP-SRK:SH2 fusio
425	7	1.6	228	20	AAW94261	CIIH6FTXN123 fusio	498	7	1.6	384	20	AAW96822	A fusion protein o
426	7	1.6	229	22	AAU16484	Human novel secret	499	7	1.6	384	23	AAU85596	Ra12-L985P fusion
427	7	1.6	230	22	AAO22139	Ra12-DPPD fusion p	500	7	1.6	386	22	AAW20280	Moraxella catarrha
428	7	1.6	231	20	AAV32071	Mycobacterium tube	501	7	1.6	388	22	ABW71779	Drosophila melanog
429	7	1.6	234	19	AAW69608	Human cysteine mod	502	7	1.6	391	21	AAV53267	Phagemid ph8 wild
430	7	1.6	234	23	AAU74600	Antigenic fusion p	503	7	1.6	391	22	ABW61179	Drosophila melanog
431	7	1.6	235	19	AAW69611	Human cysteine mod	504	7	1.6	395	23	ABW77556	His tagged cytochr
432	7	1.6	236	19	AAW69612	Human cysteine mod	505	7	1.6	400	22	AAU69907	Human prostate pro
433	7	1.6	238	22	AAU38920	C. trachomatis LI	506	7	1.6	400	22	AAW01362	Ra12-P501S-E2 cons
434	7	1.6	242	22	AAU01883	M. tuberculosis an	507	7	1.6	400	23	ABW95367	Ra12-P501S-E2 cons
435	7	1.6	249	23	ABW94257	Chlamydia protein	508	7	1.6	401	20	AAV06421	Saccharomyces cere
436	7	1.6	252	13	AAW22140	CAP-C fusion prote	509	7	1.6	402	22	ABW64529	Drosophila melanog
437	7	1.6	254	16	AAW82532	Hybrid IE beta cha	510	7	1.6	403	20	AAV03201	Amino acid sequenc
438	7	1.6	255	16	AAW82533	Hybrid IE beta cha	511	7	1.6	403	21	AAV96241	T. vaginalis homoc
439	7	1.6	265	22	AAU16053	Human novel secret	512	7	1.6	403	21	AAV76757	T. vaginalis homoc
440	7	1.6	267	23	ABW74999	Ra12/N-terminal po	513	7	1.6	403	22	AAW67240	T. vaginalis homoc
441	7	1.6	269	21	AAV71010	Chicken phospholam	514	7	1.6	405	22	AAU38919	C. trachomatis hyp
442	7	1.6	273	23	AAU85590	Lung tumour protei	515	7	1.6	408	22	AAU01886	M. tuberculosis an
443	7	1.6	281	22	AAE11943	Human CG27 (or C86	516	7	1.6	408	22	AAU01887	M. tuberculosis an
444	7	1.6	285	22	AAU59550	Propionibacterium	517	7	1.6	409	23	AAE19574	Murine herpesvirus
445	7	1.6	286	23	AAU85597	Ra12S-L985P fusion	518	7	1.6	411	23	ABW79453	Single chain class
446	7	1.6	297	23	AAU81532	Human colon cancer	519	7	1.6	412	23	ABW94254	Chlamydia protein
447	7	1.6	298	22	ABW52478	Escherichia coli p	520	7	1.6	418	23	ABW93969	Herbicidally activ
448	7	1.6	299	20	AAV32067	Mycobacterium tube	521	7	1.6	419	23	ABP38596	Staphylococcus epi

522	1.6	424	22	AAB48352	TR6-Ig fusion prot	595	7	1.6	672	13	AAR31216	penicillin binding
523	1.6	431	22	AAU55084	Propionibacterium	596	7	1.6	674	22	AAB66430	pseudomonas aerugi
524	1.6	433	20	AAU32065	Mycobacterium tube	597	7	1.6	674	22	AAB66447	pseudomonas aerugi
525	1.6	433	23	ABB94255	Chlamydia protein	598	7	1.6	676	22	AAB66448	pseudomonas aerugi
526	1.6	433	23	AAU74594	Antigenic fusion p	599	7	1.6	676	22	AAB66445	Chlamydia trachoma
527	1.6	435	23	ABG34126	Human clathrin coa	600	7	1.6	683	22	AAG83282	Chlamydia trachoma
528	1.6	445	19	AAW68391	Clostridium botuli	601	7	1.6	683	23	ABB94253	Chlamydia trachoma
529	1.6	447	21	AAAG43728	Arabidopsis thalia	602	7	1.6	685	20	AAU00915	Human serum induci
530	1.6	449	22	AAU69872	Human prostate ser	603	7	1.6	685	20	AAW88432	Disease associated
531	1.6	449	22	AAU01227	P703P and PSA fusi	604	7	1.6	685	23	ABP61474	Human NF-kB activa
532	1.6	453	23	ABB95332	Human P703P/PSA fu	605	7	1.6	691	22	AG83271	Chlamydia trachoma
533	1.6	453	18	AAW17784	FIV integrase-LexA	606	7	1.6	691	23	ABB94242	Chlamydia trachoma
534	1.6	459	16	AAW66729	Aromatic dihydrodi	607	7	1.6	700	22	AAG83279	Chlamydia trachoma
535	1.6	459	23	ABB92609	Herbicidally activ	608	7	1.6	700	23	ABB94250	Chlamydia trachoma
536	1.6	465	15	AAAG63201	Wheat germ ACCase	609	7	1.6	702	22	AAB83844	Chlamydia trachoma
537	1.6	466	22	AAW96071	Putative P. abyssi	610	7	1.6	706	22	ABB65428	Amino acid sequenc
538	1.6	476	23	AAU76912	Human CRP2-4 /IgG	611	7	1.6	710	20	AAU32066	Drosophila melanog
539	1.6	484	23	AAU76916	Human zcytorII/IgG	612	7	1.6	710	23	AAE17588	Mycobacterium tube
540	1.6	486	21	AAAG28142	Arabidopsis thalia	613	7	1.6	715	20	AAU21860	Mycobacterium spec
541	1.6	487	22	AAAG83280	Chlamydia trachoma	614	7	1.6	715	22	AAG83273	Amino acid sequenc
542	1.6	487	23	ABB94251	Chlamydia trachoma	615	7	1.6	715	22	AAG83275	Chlamydia trachoma
543	1.6	499	18	AAW17783	FIV integrase-LexA	616	7	1.6	715	23	ABB94244	Chlamydia trachoma
544	1.6	507	21	AAW00158	SCD4-SCFV(17b) HIV	617	7	1.6	715	23	ABB94246	Chlamydia trachoma
545	1.6	511	23	AAW51137	Maltose binding pr	618	7	1.6	717	21	AAG52801	Arabidopsis thalia
546	1.6	517	16	AAW84619	Wheat acetyl-Coenz	619	7	1.6	717	22	ABB66734	Drosophila melanog
547	1.6	518	22	ABB61131	Drosophila melanog	620	7	1.6	723	23	AAU74595	Antigenic fusion p
548	1.6	518	22	AAAG83276	Chlamydia trachoma	621	7	1.6	726	23	AAU74588	Antigenic fusion p
549	1.6	518	23	ABB94247	Chlamydia trachoma	622	7	1.6	729	20	AAU32059	Mycobacterium tube
550	1.6	521	21	AAW54022	Human pancreatic c	623	7	1.6	729	22	AAO22142	Ra12-H9-32A fusion
551	1.6	525	21	AAW13645	C. pneumoniae sero	624	7	1.6	729	23	AAE17572	Mycobacterium spec
552	1.6	525	22	AAAG83213	Protein encoded by	625	7	1.6	729	23	AAE17573	Mycobacterium spec
553	1.6	525	23	ABB94184	Chlamydia protein	626	7	1.6	744	22	AAU01902	M. tuberculosis an
554	1.6	529	23	AAU74387	Breast tumour-spec	627	7	1.6	753	23	ABP41992	Human ovarian anti
555	1.6	543	22	AAU01905	M. tuberculosis an	628	7	1.6	768	23	ABB49385	Listeria monocytog
556	1.6	543	22	AAW66429	P. aeruginosa WpM	629	7	1.6	778	21	AAW13642	C. trachomatis pmp
557	1.6	543	22	AAW66446	Pseudomonas aerugi	630	7	1.6	778	22	AAW83210	Protein encoded by
558	1.6	553	22	ABB59793	Drosophila melanog	631	7	1.6	778	23	ABB94181	Chlamydia protein
559	1.6	554	17	AAW09252	K. pneumoniae diol	632	7	1.6	780	18	AAW11819	Candida albicans t
560	1.6	577	22	AAO22140	Ra12-WT1 fusion pr	633	7	1.6	785	23	ABP30430	Streptococcus poly
561	1.6	580	21	AAW82157	Pseudomonas putida	634	7	1.6	788	22	AAU01903	M. tuberculosis an
562	1.6	583	22	AAW83281	Chlamydia trachoma	635	7	1.6	815	22	AAU01904	M. tuberculosis an
563	1.6	583	23	ABB94252	Chlamydia trachoma	636	7	1.6	816	22	AAU03628	Group B Streptococ
564	1.6	585	22	AAW83277	Chlamydia trachoma	637	7	1.6	816	23	ABP25423	Streptococcus poly
565	1.6	585	23	ABB94248	Chlamydia trachoma	638	7	1.6	821	18	AAW16611	Candida albicans t
566	1.6	596	20	AAU32070	Mycobacterium tube	639	7	1.6	821	21	AAW13644	C. trachomatis pmp
567	1.6	596	22	AAU30320	Novel human secret	640	7	1.6	821	22	AAG83212	Protein encoded by
568	1.6	596	23	AAE17574	Mycobacterium spec	641	7	1.6	821	23	ABB94183	Chlamydia protein
569	1.6	597	22	AAU08231	Polypeptide encode	642	7	1.6	824	11	AAW04575	Derived amino acid
570	1.6	598	22	AAU38922	C. trachomatis CT6	643	7	1.6	825	12	AAW11254	Human IL-4 recepto
571	1.6	599	23	AAU74599	Antigenic fusion p	644	7	1.6	825	18	AAW13499	Human interleukin-
572	1.6	600	20	AAU32068	Mycobacterium tube	645	7	1.6	825	19	AAW60668	Human interleukin-
573	1.6	600	23	AAU74597	Antigenic fusion p	646	7	1.6	825	19	AAW48314	Human interleukin-
574	1.6	605	20	AAU21870	Amino acid sequenc	647	7	1.6	825	20	AAW93004	Human IL-4 recepto
575	1.6	609	20	AAU41279	Fusion protein con	648	7	1.6	825	20	AAW73470	Human interleukin-
576	1.6	619	22	AAW83270	Chlamydia trachoma	649	7	1.6	825	22	AAW38752	Human polypeptide
577	1.6	619	23	ABB94241	Chlamydia trachoma	650	7	1.6	825	22	AAW66970	Human IL4Ralpha
578	1.6	631	22	AAW83274	Chlamydia trachoma	651	7	1.6	825	22	AAU07612	Human interleukin-
579	1.6	631	23	ABB94245	Chlamydia trachoma	652	7	1.6	825	23	AAU77212	Human interleukin-
580	1.6	638	22	ABB58483	Drosophila melanog	653	7	1.6	825	23	AAU77214	Human interleukin-
581	1.6	646	20	AAW17892	HCV NS4A-NS3 compl	654	7	1.6	825	23	AAU78090	Human interleukin-
582	1.6	646	20	AAW17894	HCV NS4A-NS3 compl	655	7	1.6	825	23	AAW07427	Human interleukin-
583	1.6	646	20	AAU24950	HCV NS4A-NS3 compl	656	7	1.6	826	21	AAW95050	Candida albicans p
584	1.6	646	22	AAW83272	Chlamydia trachoma	657	7	1.6	826	22	AAW20493	Human interleukin-
585	1.6	646	23	ABB94243	Chlamydia trachoma	658	7	1.6	831	23	AAU77217	Human interleukin-
586	1.6	654	20	AAU21859	Amino acid sequenc	659	7	1.6	840	21	AAW93652	A mammalian solubl
587	1.6	654	22	AAW83278	Chlamydia trachoma	660	7	1.6	840	21	AAW70028	Soluble Interleuki
588	1.6	654	23	ABB94249	Chlamydia trachoma	661	7	1.6	856	20	AAW32064	Mycobacterium tube
589	1.6	660	22	AAU38921	C. trachomatis CT8	662	7	1.6	859	23	AAU74593	Antigenic fusion p
590	1.6	663	21	AAW52602	Arabidopsis thalia	663	7	1.6	861	22	ABB62534	Drosophila melanog
591	1.6	666	20	AAU21871	Amino acid sequenc	664	7	1.6	869	15	AAW56501	TATA-binding prote
592	1.6	666	22	ABB71578	Drosophila melanog	665	7	1.6	869	17	AAW06091	Human TATA-binding
593	1.6	667	20	AAW17893	HCV NS4A-NS3 compl	666	7	1.6	879	18	AAW25022	TATA-binding prote
594	1.6	667	20	AAW17891	HCV NS4A-NS3 compl	667	7	1.6	879	22	ABW25755	Novel human diagno

668	7	1.6	879	22	ABG28383	Novel human diaphanous	741	6	1.4	6	21	AA995641	His6 tag used to purify
669	7	1.6	888	23	AB91325	Herbicidally active	742	6	1.4	6	21	AA987720	Thrombin cleavage
670	7	1.6	948	21	AAB13643	C. trachomatis pmp	743	6	1.4	6	21	AA996204	Thrombin cleavage
671	7	1.6	948	22	AAG83211	Protein encoded by	744	6	1.4	6	21	AA990140	Thrombin cleavage
672	7	1.6	948	23	AB94182	Chlamydia protein	745	6	1.4	6	21	AA990451	Thrombin cleavage
673	7	1.6	950	21	AAG41703	Arabidopsis thaliana	746	6	1.4	6	21	AA981407	Thrombin cleavage
674	7	1.6	960	21	AAG31324	Arabidopsis thaliana	747	6	1.4	6	21	AA981407	Thrombin cleavage
675	7	1.6	982	23	AAU71855	Leishmania antigen	748	6	1.4	6	21	AA951517	Thrombin cleavage
676	7	1.6	989	23	AB92703	Leishmania antigen	749	6	1.4	6	21	AA977706	Thrombin cleavage
677	7	1.6	1024	22	AB961798	Herbicidally active	750	6	1.4	6	21	AA978346	His-6 peptide epit
678	7	1.6	1068	20	AA930048	Drosophila melanog	751	6	1.4	6	21	AA978351	His-6 peptide SQ
679	7	1.6	1079	22	AA974830	Pancreatic eukaryo	752	6	1.4	6	22	AB976820	His tag.
680	7	1.6	1115	20	AA930047	Prostate tumour an	753	6	1.4	6	22	AA952173	His tag. Unidenti
681	7	1.6	1115	23	AAU81285	Pancreatic eukaryo	754	6	1.4	6	22	AA952173	Peptide tag #1. Sy
682	7	1.6	1161	22	AB970419	EIF-2alpha kinase.	755	6	1.4	6	22	AA951707	Peptide tag #7 use
683	7	1.6	1186	22	AB963516	Drosophila melanog	756	6	1.4	6	22	AA951720	FSH alpha or beta
684	7	1.6	1225	22	ABG24444	Drosophila melanog	757	6	1.4	6	22	AA951720	HexaHis tag used t
685	7	1.6	1255	20	AA905944	Novel human diagno	758	6	1.4	6	22	AA956669	His tag used for r
686	7	1.6	1284	22	AA971869	Thermophilus therm	759	6	1.4	6	22	AA956669	Poly-His tag, to g
687	7	1.6	1284	23	AB908376	Drosophila melanog	760	6	1.4	6	22	AA956669	6-His epitope tag
688	7	1.6	1317	23	AA936781	D. melanogaster CG	761	6	1.4	6	22	AA956669	Amino acid sequenc
689	7	1.6	1427	23	AAU71857	A3 maize ACCase.	762	6	1.4	6	22	AA956669	His tag useful as
690	7	1.6	1465	22	AB964500	Leishmania MAPS1A-	763	6	1.4	6	22	AA956669	Metal capturing pr
691	7	1.6	1534	22	AB958402	Drosophila melanog	764	6	1.4	6	22	AA956669	His-tag peptide.
692	7	1.6	1561	22	AB966445	Drosophila melanog	765	6	1.4	6	22	AA956669	6-His epitope used
693	7	1.6	1641	23	AAU71856	Drosophila melanog	766	6	1.4	6	22	AA956669	C-terminal His tag
694	7	1.6	1807	22	AA985697	Leishmania antigen	767	6	1.4	6	22	AA956669	Poly-His tag used
695	7	1.6	1895	22	AB970088	Recombinant protei	768	6	1.4	6	22	AA956669	MSA1 gp42 C-termin
696	7	1.6	1911	22	AB961038	Drosophila melanog	769	6	1.4	6	22	AA956669	His Tag for purifi
697	7	1.6	2028	22	AB985698	Drosophila melanog	770	6	1.4	6	22	AA956669	Nascent protein de
698	7	1.6	2091	22	AB970224	Recombinant protei	771	6	1.4	6	22	AA956669	6-His epitope. Un
699	7	1.6	2280	22	AB961650	Drosophila melanog	772	6	1.4	6	22	AA956669	Transdominant effe
700	7	1.6	2311	23	AAU11820	Drosophila melanog	773	6	1.4	6	22	AA956669	6-His peptide epit
701	7	1.6	2325	17	AAW05590	Wheat plastid acet	774	6	1.4	6	22	AA956669	

814	6	1.4	8	23	AAU99814	Membrane associate	887	11	23	ABB76132	Modified polyhisti
815	6	1.4	8	23	AAU96775	Divinyl ether synt	888	11	23	AAE20416	His6Gly4Cys spacer
816	6	1.4	8	23	AAU78663	His tag sequence.	889	11	23	AAM50625	Polyhistidine tag
817	6	1.4	8	23	AAU77821	C terminal extensio	890	12	11	AAR07064	Transcript of plas
818	6	1.4	8	23	AAU77832	His tag peptide.	891	12	15	AAR60140	PGEX-27k protease
819	6	1.4	8	23	AAE16245	Polyhistidine tag,	892	12	16	AAR66449	GST-hippuricase fu
820	6	1.4	9	17	AAW05142	Variant adenovirus	893	12	18	AAW10571	Factor Xa recognit
821	6	1.4	9	17	AAU97378	Penta-histidine pe	894	12	19	AAW61549	Endoprotease Xa
822	6	1.4	9	19	AAW60080	Homo sapiens RH ep	895	12	20	AAV16775	Histidine tag used
823	6	1.4	9	20	AAU29934	FLAG peptide. Syn	896	12	20	AAW94420	Plasmodium falcipa
824	6	1.4	9	20	AAU28306	Amino acid sequenc	897	12	20	AAW82993	Human fibronectin
825	6	1.4	9	20	AAW90199	B7.2-6His tag fusi	898	12	21	AAU95878	Hexahistidine tag
826	6	1.4	9	20	AAW82990	Human fibronectin	899	12	22	AAAG64365	Peptide insert #5.
827	6	1.4	9	21	AAU15187	His-containing pep	900	12	22	AAAB85050	Peptide tag for re
828	6	1.4	9	21	AAU79558	Peptide tag 6xHis.	901	12	22	AAAB83199	FLAG epitope tag.
829	6	1.4	9	21	AAU53258	Human neurotactin	902	12	22	AAAB72438	His tag. Unidenti
830	6	1.4	9	21	AAU44463	Gly(His)8 tag, gh	903	12	23	AAU96774	Divinyl ether synt
831	6	1.4	9	22	AAE13114	C-terminal tag of	904	12	23	AAU96716	Molecular marker f
832	6	1.4	9	22	AAAB82692	His-tag. Syntheti	905	12	23	AAE19655	Arabidopsis thalia
833	6	1.4	9	22	AAAG64361	Peptide insert #1.	906	12	23	ABB05983	Monoclonal antibod
834	6	1.4	9	22	AAAB99760	Rhesus D antibody	907	12	16	AAAB72792	Hexa-histidine lea
835	6	1.4	9	22	AAU97661	Influenza haemaggl	908	13	19	AAW69482	Partial sequence o
836	6	1.4	9	23	AAU09999	Nonapeptide encodi	909	13	19	AAW61548	Endoprotease Xa
837	6	1.4	10	14	AAU39733	First type QB- C-t	910	13	20	AAU17524	P. carinii tyrosyl
838	6	1.4	10	15	AAU73684	Metal-affinity dec	911	13	20	AAU05816	Nq-toxic modified
839	6	1.4	10	16	AAU74672	Polyhistidine moti	912	13	21	AAU10782	B. subtilis lumazi
840	6	1.4	10	16	AAU77469	I-domain N-termina	913	13	21	AAU14343	Coriolus versicolo
841	6	1.4	10	18	AAW31496	Human DNase II pro	914	13	21	AAU56262	Maedi-Visna virus
842	6	1.4	10	18	AAW30084	Human calcitonin g	915	13	22	AAE13116	C-terminal tag of
843	6	1.4	10	19	AAU72331	Protein allergen I	916	13	22	AAU09362	N-terminal HMK-His
844	6	1.4	10	19	AAW56089	Human monocyte che	917	13	22	AAU09363	N-terminal HMK-His
845	6	1.4	10	20	AAU43534	Leishmania mexican	918	13	22	AAU09364	N-terminal HMK-His
846	6	1.4	10	20	AAU50501	Dermatophagoides s	919	13	22	AAE12152	Oligopeptide used
847	6	1.4	10	20	AAU06610	Hexahistidine tag	920	13	22	AAAG0817	His tag peptide #1
848	6	1.4	10	20	AAW94259	His-tag sequence.	921	13	22	AAAG0820	His tag peptide #2
849	6	1.4	10	20	AAW89262	N-terminal hexahis	922	13	22	AAAB72415	Synthetic leader s
850	6	1.4	10	20	AAW82996	Human fibronectin	923	13	22	AAAB52258	pCONGAH His tag pe
851	6	1.4	10	21	AAU10776	B. subtilis lumazi	924	13	22	AAU11796	Peptide adapter fo
852	6	1.4	10	21	AAU10830	Expression vector	925	13	23	AAU11045	Translated linker
853	6	1.4	10	21	AAU28976	N-terminal of pept	926	13	14	AAU32952	Mastoparan analogu
854	6	1.4	10	21	AAU87721	Leader sequence of	927	14	14	AAU34641	Mastoparan mutant
855	6	1.4	10	21	AAU90141	Leader sequence pe	928	14	14	AAU60018	Can fi peptide wit
856	6	1.4	10	21	AAU51518	Human TRFP constru	929	14	15	AAU60018	Scrv expression ca
857	6	1.4	10	22	AAW43368	Mycoplasma genital	930	14	16	AAU64816	Chimeric Tag DNA p
858	6	1.4	10	22	AAW43369	Mycoplasma genital	931	14	21	AAU33455	Peptide #1. Unide
859	6	1.4	10	22	AAW43480	Mycoplasma genital	932	14	21	AAU28977	Coiled coil protei
860	6	1.4	10	22	AAW43481	Mycoplasma genital	933	14	21	AAU10559	Feline TRFP protei
861	6	1.4	10	22	AAW70426	Oligo-histidine pu	934	14	21	AAU87722	Synthetic peptide
862	6	1.4	10	22	AAU72220	Histidine purifica	935	14	21	AAU98160	Leader sequence pe
863	6	1.4	10	22	AAU97658	Influenza haemaggl	936	14	21	AAU51519	Human TRFP derived
864	6	1.4	10	22	AAU19104	His-Tag containing	937	14	22	AAU98084	Human peptide #135
865	6	1.4	10	23	AAU97726	Epitope tag #2 use	938	14	22	AAU09361	N-terminal HMK-His
866	6	1.4	10	23	ABB76374	Hexa-histidine tag	939	14	22	AAE10674	Peptide useful to
867	6	1.4	10	23	AAU78933	9 Histidine peptid	940	14	22	AAE12159	Human hG-CSF N-ter
868	6	1.4	10	23	AAU14437	Peptide derived fr	941	14	22	AAAB82693	His-tag. Syntheti
869	6	1.4	10	23	ABB05988	Monoclonal antibod	942	14	22	AAAB73439	C-terminally His-t
870	6	1.4	11	14	AAU36558	Sequence surroundi	943	14	22	AAE92238	Toxin peptide SEQ
871	6	1.4	11	16	AAU77410	Extension peptide	944	14	22	AAE00403	GST-PROST 07 fusio
872	6	1.4	11	19	AAW61480	A. fumigatus aller	945	14	23	ABB08199	Amino acid sequenc
873	6	1.4	11	19	AAW53171	C-terminal peptide	946	14	23	ABB05986	Monoclonal antibod
874	6	1.4	11	20	AAU42688	Polyhistidine tag	947	14	23	AAAG80199	Human cystatin C d
875	6	1.4	11	20	AAU05815	Non-toxic modified	948	14	15	AAAG60513	Hexahistidine-cont
876	6	1.4	11	20	AAW80426	Linker used in the	949	15	21	AAU96095	Hexahistidine tag.
877	6	1.4	11	20	AAW70581	Amino acid sequenc	950	15	21	AAU84236	Phosphorylation si
878	6	1.4	11	21	AAU09993	C. tetani tetanus	951	15	21	AAU69625	Junction region of
879	6	1.4	11	21	AAU81810	Affinity peptide t	952	15	21	AAU69626	Junction region of
880	6	1.4	11	22	AAU68075	Human Breast cance	953	15	21	AAU69627	Junction region of
881	6	1.4	11	22	AAU68080	Human Breast cance	954	15	22	AAU52533	His tag peptide.
882	6	1.4	11	22	AAE11753	Modified polyhisti	955	15	22	AAE13115	C-terminal tag of
883	6	1.4	11	22	AAU99770	Rhesus D antibody	956	15	22	AAU79094	Amino acid sequenc
884	6	1.4	11	22	AAU68981	Fluorescent isoele	957	15	22	AAE08870	Escherichia coli c
885	6	1.4	11	23	AAU97110	pAMG21-human MK61	958	15	22	AAAG64156	Recombinant HCV NS
886	6	1.4	11	23	AAU96727	Molecular marker f	959	15	22	AAAB67388	N-terminal His tag

960 6 1.4 15 22 AAB66370 Hepatitis C virus  
961 6 1.4 15 23 AAE20475 Human pancreatic r  
962 6 1.4 15 23 AAE19656 Arabidopsis thalia  
963 6 1.4 15 23 AAU77664 His tag peptide.  
964 6 1.4 15 23 AAB05375 15 amino acid His  
965 6 1.4 16 14 AAR44552 Feline leukaemia v  
966 6 1.4 16 16 AAW11923 Residues 146-161 o  
967 6 1.4 16 16 AAR69813 Cleavage site reco  
968 6 1.4 16 18 AAW80121 Plasmodium falcipa  
969 6 1.4 16 20 AAY06321 His-c-myc epitope  
970 6 1.4 16 21 AAY51161 Modified VPI pepti  
971 6 1.4 16 22 AAU27552 Tag sequence used  
972 6 1.4 16 22 AAY72938 N-terminal of huma  
973 6 1.4 16 22 AAY97659 Influenza haemaggl  
974 6 1.4 16 22 AAB66989 Peptide: SEQ ID 16  
975 6 1.4 16 23 ABB08039 Amino acid fragmen  
976 6 1.4 16 23 AAU77202 Polyasparagine, he  
977 6 1.4 16 23 AAM47310 GMT-JMT fusion pro  
978 6 1.4 17 13 AAR25383 Junction sequence  
979 6 1.4 17 14 AAR37708 Delta14 Ser17 hcNT  
980 6 1.4 17 17 AAR97389 CC49 VH-spacer-PLA  
981 6 1.4 17 21 AAB26262 N-terminal fusion  
982 6 1.4 17 22 AAG99608 ERA binding domain  
983 6 1.4 18 15 AAR60253 Vector-encoded NF-  
984 6 1.4 18 16 AAR78152 Synthetic leader t  
985 6 1.4 18 16 AAR69814 Site recognised En  
986 6 1.4 18 17 AAR95870 CR2 receptor ligan  
987 6 1.4 18 21 AAB11047 Staphylococcal sur  
988 6 1.4 18 21 AAB10856 pSectag-VEGF prote  
989 6 1.4 18 22 ABB28438 Peptide #1089 enco  
990 6 1.4 18 22 ABB31815 Peptide #4466 enco  
991 6 1.4 18 22 ABB33620 Peptide #1126 enco  
992 6 1.4 18 22 ABB37046 Peptide #4552 enco  
993 6 1.4 18 22 ABB19075 Protein #1074 enco  
994 6 1.4 18 22 ABB22361 Protein #4360 enco  
995 6 1.4 18 22 AAB85750 Mouse sonic hedgeh  
996 6 1.4 18 22 AAM54393 Human brain expres  
997 6 1.4 18 22 AAM57773 Human brain expres  
998 6 1.4 18 22 AAM66794 Human bone marrow  
999 6 1.4 18 22 AAM70188 Human bone marrow  
1000 6 1.4 18 22 AAM14660 Peptide #1094 enco

ALIGNMENTS

RESULT 1  
AAB83926  
ID AAB83926 standard; Protein; 375 AA.  
XX AAB83926;  
AC (first entry)  
DT 23-JUL-2001  
DE A major merozoite surface protein-1 fragment of 42kDa.  
XX Major merozoite surface protein-1; MSP1-42; melittin signal peptide;  
KW malaria vaccine.  
XX Plasmodium falciparum.  
OS Plasmodium falciparum.  
PN WC2001134188-A1  
XX 17-MAY-2001.  
XX 09-NOV-2000; 2000WO-US31064.  
XX 12-NOV-1999; 99US-0165178.  
PR 01-DEC-1999; 99US-0168327.  
PR 22-AUG-2000; 2000US-0226861.  
XX (UYHA-) UNIV HAWAII.  
PA (UYCH-) UNIV CHINESE HONG KONG.

NEW  
102(e)

PA (QUEE-) QUEEN EMMA FOUND.  
XX Hui GSN, Lap-Yin P, Ho WKK;  
XX WPI; 2001-335879/35.  
DR N-PSDB; AAF89840.  
XX Producing malaria vaccine, useful for treatment or prevention of all  
forms of malaria in humans, by expressing immunogenic merozoite protein  
fragment in a baculovirus system  
XX Example 3; Page 87-88; 95pp; English.  
XX The present sequence represents a major merozoite surface protein-1  
C-terminal fragment of 42kDa (MSP1-42). This fragment is linked to a  
melittin signal peptide, and then expressed in a in a  
silkworm/baculovirus system. The protein is used to prepare a  
malaria vaccine, which is used to treat or prevent malaria, caused by  
any of the four species of Plasmodium that infect humans.  
XX Sequence 375 AA;

Query Match 87.0%; Score 375; DB 22; Length 375;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 57 MAISVTMDNILSGFENEYDVIYKPLAGVYRSLKQIEKNFTFNLDILNSRLKRRK 116  
Db 1 MAISVTMDNILSGFENEYDVIYKPLAGVYRSLKQIEKNFTFNLDILNSRLKRRK 60  
Qy 117 YFLOVLESDLAQFHRHISSEYIIEEDSFLLNSEQNTLLSKYIKESVENDIKFAQEGI 176  
Db 61 YFLOVLESDLAQFHRHISSEYIIEEDSFLLNSEQNTLLSKYIKESVENDIKFAQEGI 120  
Qy 177 SYEYKVLAKYKDDLESIKKVIKEEKEPPSPPTPPSPAKTDEOKKESKFLPFLTNLET 236  
Db 121 SYEYKVLAKYKDDLESIKKVIKEEKEPPSPPTPPSPAKTDEOKKESKFLPFLTNLET 180  
Qy 237 LYNNLVNKIDYDYLNLKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFKNPYDFAIK 296  
Db 181 LYNNLVNKIDYDYLNLKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFKNPYDFAIK 240  
Qy 297 KLINDTKKMDLGLSTGLVONFPNTIISKLEGGKFDMDLNIHQCVKQCPENSGCF 356  
Db 241 KLINDTKKMDLGLSTGLVONFPNTIISKLEGGKFDMDLNIHQCVKQCPENSGCF 300  
Qy 357 RHLDEREECKLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSRKKITCECT 416  
Db 301 RHLDEREECKLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSRKKITCECT 360  
Qy 417 KPDSYPLFDGIFCSS 431  
Db 361 KPDSYPLFDGIFCSS 375

RESULT 2  
AAW36103  
ID AAW36103 standard; Protein; 116 AA.  
XX AAW36103;  
AC (first entry)  
DT 25-MAR-1998  
DE PfMSP1(p19)A protein sequence.  
XX Plasmodium vivax; merozoite surface protein; MSP1; p19;  
KW Plasmodium falciparum; malaria; vaccine; immunity; epitope.  
XX Plasmodium falciparum.  
OS Synthetic.  
XX Key Location/Qualifiers  
FT Region 1..95



```
FT FT /note= "amino acids derived from P. falciparum MSP1 p19  
FT FT fragment"  
FT FT 96..116  
XX /note= "glycosylphosphatidylinositol anchoring sequence"  
XX WO9730158-A2.  
XX 21-AUG-1997.  
XX (INSP ) INST PASTEUR.  
XX (UJNY ) UNIV NEW YORK STATE.  
XX 14-FEB-1997; 97WO-FR00290.  
XX 14-FEB-1996; 96FR-0001822.  
XX (INSP ) INST PASTEUR.  
XX (UJNY ) UNIV NEW YORK STATE.  
XX Barnwell JW, Longacre-Andre S, Mendis K, Nato F;  
XX Roth C;  
XX WPI; 1997-425033/39.  
XX N-PSDB; AAT94550.  
XX Recombinant protein containing the merozoite surface protein-1 p19  
FT fragment - useful in anti-malarial vaccines, diagnosis and protein  
FT purification  
XX Disclosure; Fig 1B; 85pp; French.  
XX This is the amino acid sequence of a recombinant protein comprising  
CC amino acids 1613-1705 of the Plasmodium falciparum merozoite surface  
CC protein 1 (MSP1) 19 kD C-terminal fragment (p19), linked to a  
CC glycosylphosphatidylinositol membrane anchoring sequence. p19 is the  
CC C-terminal fragment of the 42 kD MSP1 from Plasmodium species.  
CC The recombinant protein can be used for the production of anti-malarial  
CC vaccines, where the p19 fragment provides a high level of protective  
CC immunity since it includes epitopes not presented in the p42 fragment.  
XX  
SQ Sequence 116 AA:  
Query Match 13.9%; Score 60; DB 18; Length 116;  
Best Local Similarity 100.0%; Pred. No. 3.7e-52;  
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 338 NISQHCYKVKQCPENSGCFRHLDERECKLLNKKQEGDKCVENPNTCNENNGGDADA 397  
DB 3 NISQHCYKVKQCPENSGCFRHLDERECKLLNKKQEGDKCVENPNTCNENNGGDADA 62  
RESULT 3  
AAW22592  
ID AAW22592 standard; Protein; 116 AA.  
XX AC AAW22592;  
XX 25-MAR-1998 (first entry)  
XX PfMSPl(p19)A protein sequence.  
XX Plasmodium vivax; merozoite surface protein; MSP1; p19;  
KW Plasmodium falciparum; malaria; vaccine; immunity; epitope.  
XX Plasmodium falciparum.  
OS Synthetic.  
XX Key Location/Qualifiers  
XX FT Region 1..95  
FT /note= "amino acids derived from P. falciparum MSP1 p19  
FT fragment"  
FT Region 96..116  
FT /note= "glycosylphosphatidylinositol anchoring sequence"  
XX WO9730159-A2.  
XX
```

```
PD 21-AUG-1997.  
XX 14-FEB-1997; 97WO-FR00291.  
XX 14-FEB-1996; 96FR-0001821.  
XX (INSP ) INST PASTEUR.  
XX (UJNY ) UNIV NEW YORK STATE.  
XX Barnwell JW, Longacre-Andre S, Mendis K, Nato F;  
XX Roth C;  
XX WPI; 1997-425034/39.  
XX P-PSDB; AAW22592.  
XX Recombinant protein containing Plasmodium merozoite surface.  
FT protein-1 p42 fragment - useful in antimalarial vaccines, also new  
FT antibodies for diagnosis and protein purification  
XX Disclosure; Fig 1B; 85pp; French.  
XX This is the amino acid sequence of a recombinant protein comprising  
CC amino acids 1613-1705 of the Plasmodium falciparum merozoite surface  
CC protein 1 (MSP1) 19 kD C-terminal fragment (p19), linked to a  
CC glycosylphosphatidylinositol membrane anchoring sequence. p19 is the  
CC C-terminal fragment of the 42 kD MSP1 from Plasmodium species.  
CC The recombinant protein can be used for the production of anti-malarial  
CC vaccines, where the p19 fragment provides a high level of protective  
CC immunity since it includes epitopes not presented in the p42 fragment.  
XX  
SQ Sequence 116 AA:  
Query Match 13.9%; Score 60; DB 18; Length 116;  
Best Local Similarity 100.0%; Pred. No. 3.7e-52;  
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 338 NISQHCYKVKQCPENSGCFRHLDERECKLLNKKQEGDKCVENPNTCNENNGGDADA 397  
DB 3 NISQHCYKVKQCPENSGCFRHLDERECKLLNKKQEGDKCVENPNTCNENNGGDADA 62  
RESULT 4  
AAW22593  
ID AAW22593 standard; Protein; 127 AA.  
XX AC AAW22593;  
XX 25-MAR-1998 (first entry)  
XX PfMSPl(p19)S protein sequence.  
XX Chimeric; Plasmodium vivax; merozoite surface protein; MSP1; p19;  
KW Plasmodium falciparum; malaria; vaccine; immunity; epitope.  
XX Chimeric - Plasmodium vivax.  
OS Chimeric - Plasmodium falciparum.  
XX Key Location/Qualifiers  
XX FT Peptide 1..19  
FT /note= "signal peptide"  
FT Protein 20..127  
FT /note= "mature protein"  
FT Region 1..32  
FT /note= "derived from P. vivax MSP1"  
FT Region 33..34  
FT /note= "encoded by restriction enzyme sequence used to  
FT create the chimeric sequence"  
FT Region 35..127  
FT /note= "derived from P. falciparum C-terminal p19  
FT fragment of MSP1"  
XX WO9730159-A2.  
XX
```

```

PD 21-AUG-1997.
XX
XX 14-FEB-1997; 97WO-FR00291.
XX
XX 14-FEB-1996; 96FR-0001821.
XX
XX (INSP ) INST PASTEUR.
PA (UYN ) UNIV NEW YORK STATE.
XX
XX Barnwell JW, Longacre-Andre S, Mendis K, Nato F;
PI Roth C;
XX
XX WPI; 1997-425034/39.
DR P-PSDB; AAW22592.
XX
XX Recombinant protein containing Plasmodium merozoite surface
PT protein-1 p42 fragment - useful in antimalarial vaccines, also new
PT antibodies for diagnosis and protein purification
XX
XX Disclosure; Fig 1C; 85pp; French.
XX
XX This is the amino acid sequence of a chimeric protein comprising amino
CC acids 1-32 of the Plasmodium vivax merozoite surface protein 1 (MSp1)
CC linked to amino acids Asn1613-Ser1705 which correspond to the 19 kD
CC C-terminal fragment (p19) of MSP-1 from Plasmodium falciparum.
CC p19 is the C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
CC The recombinant protein can be used for the production of anti-malarial
CC vaccines, where the p19 fragment provides a high level of protective
CC immunity since it includes epitopes not presented in the p42 fragment.
XX
XX Sequence 127 AA;

Query Match 13.9%; Score 60; DB 18; Length 127;
Best Local Similarity 100.0%; Pred. No. 4e-52;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 338 NISQHCVKKQCPNSGCFRHLDERECKLLNYKQBGDKCVENPNTCTNENGGCDADA 397
DB 35 NISQHCVKKQCPNSGCFRHLDERECKLLNYKQBGDKCVENPNTCTNENGGCDADA 94

RESULT 5
AAW36102
ID AAW36102 standard; Protein; 127 AA.
XX
XX AAW36102;
XX
XX 25-MAR-1998 (first entry)
XX
XX PfMSP1(p19)S protein sequence.
XX
XX Chimeric; Plasmodium vivax; merozoite surface protein; MSP1; p19;
KW Plasmodium falciparum; malaria; vaccine; immunity; epitope.
XX
XX Chimeric - Plasmodium vivax.
OS Chimeric - Plasmodium falciparum.
XX
XX Key Location/Qualifiers
FH Peptide 1..19 /note= "signal peptide"
FT Protein 20..127 /note= "mature protein"
FT Region 1..32 /note= "derived from P. vivax MSP1"
FT Region 33..34 /note= "encoded by restriction enzyme sequence used to
create the chimeric sequence"
FT Region 35..127 /note= "derived from P. falciparum C-terminal p19
fragment of MSP1"
XX
XX WO9730158-A2.
XX

```

```

PD 21-AUG-1997.
XX
XX 14-FEB-1997; 97WO-FR00290.
XX
XX 14-FEB-1996; 96FR-0001822.
XX
XX (INSP ) INST PASTEUR.
PA (UYN ) UNIV NEW YORK STATE.
XX
XX Barnwell JW, Longacre-Andre S, Mendis K, Nato F;
PI Roth C;
XX
XX WPI; 1997-425033/39.
DR N-PSDB; AAT94549.
XX
XX Recombinant protein containing the merozoite surface protein-1 p19
PT fragment - useful in anti-malarial vaccines, diagnosis and protein
PT purification
XX
XX Disclosure; Fig 1C; 85pp; French.
XX
XX This is the amino acid sequence of a chimeric protein comprising amino
CC acids 1-32 of the Plasmodium vivax merozoite surface protein 1 (MSp1)
CC linked to amino acids Asn1613-Ser1705 which correspond to the 19 kD
CC C-terminal fragment (p19) of MSP-1 from Plasmodium falciparum.
CC p19 is the C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
CC The recombinant protein can be used for the production of anti-malarial
CC vaccines, where the p19 fragment provides a high level of protective
CC immunity since it includes epitopes not presented in the p42 fragment.
XX
XX Sequence 127 AA;

Query Match 13.9%; Score 60; DB 18; Length 127;
Best Local Similarity 100.0%; Pred. No. 4e-52;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 338 NISQHCVKKQCPNSGCFRHLDERECKLLNYKQBGDKCVENPNTCTNENGGCDADA 397
DB 35 NISQHCVKKQCPNSGCFRHLDERECKLLNYKQBGDKCVENPNTCTNENGGCDADA 94

RESULT 6
AAR41357
ID AAR41357 standard; peptide; 54 AA.
XX
XX AAR41357;
XX
XX 04-MAR-1994 (first entry)
XX
XX MSP1EGF2B EGF2-like domain variant.
XX
XX Epidermal growth factor 1; merozoite surface protein 1; malaria;
KW vaccine.
XX
XX Plasmodium yoelii.
OS
XX Key Location/Qualifiers
FH Cleavage-site 1 /note= "introduced to facilitate cleavage
from recombinant protein"
FT
XX WO9317107-A.
XX
XX 02-SEP-1993.
XX
XX 22-FEB-1993; 93WO-GB00367.
XX
XX 22-FEB-1992; 92GB-0003821.
XX
XX (MEDI-) MEDICAL RES COUNCIL.
XX
XX Blackman MJ, Chappel JA, Holder AA;
XX

```

DR WPI: 1993-288413/36.  
XX Allelic variants of epidermal growth factor 1- or 2-like domains - of  
PT merozoite surface protein 1, produced recombinantly for malaria  
PT vaccines  
XX  
XX Claim 2; Fig 2b; 35pp: English.  
XX  
XX The sequence is that of an allelic variant of a merozoite surface  
CC protein-1 epidermal growth factor (EGF) 2-like domain. It may be  
CC used alone or as part of a fusion protein of EGF-1-like and  
CC EGF-2-like domains in vaccines against malaria. When expressed  
CC recombinantly it is produced in a form indistinguishable from  
CC that in the native protein.  
XX  
XX  
SQ Sequence 54 AA;  
Query Match 11.8%; Score 51; DB 14; Length 54;  
Best Local Similarity 100.0%; Pred. No. 2.1e-43;  
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 381 NPNTCNENGGCDATCEEDSGSRKKITCECTKPDSPYPLFDGIFCSS 431  
DB 2 NPNTCNENGGCDATCEEDSGSRKKITCECTKPDSPYPLFDGIFCSS 52  
RESULT 7  
AAB26110  
ID AAB26110 standard; Protein; 76 AA.  
XX  
AC AAB26110;  
XX  
XX 30-JAN-2001 (first entry)  
XX  
XX H. contortus clone-65e vector pET30a cloning junction #2.  
XX  
XX Nematode; parasite; helminth; sheep; goat; stomach; vaccine.  
XX  
XX Synthetic.  
XX  
XX WO200056763-A1.  
XX  
XX 28-SEP-2000.  
XX  
XX 16-MAR-2000; 2000WO-AU00210.  
XX  
XX 18-MAR-1999; 99AU-0009297.  
XX  
XX (NOVS) NOVARTIS AG.  
XX  
XX (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.  
XX  
XX Savin KW, Cook VR, Chen Y, Sexton JL, Apos E, Wilson LR;  
PI Griffiths TM, Newton SE;  
XX  
XX WPI: 2000-594573/56.  
DR N-PSDB; AAA94064.  
XX  
XX New Haemonchus contortus polypeptide for inducing a protective effect  
PT against a helminth by controlling helminth infection, growth, viability  
PT and/or egg fecundity and for ameliorating the symptoms of helminth  
PT infection -  
XX  
XX Disclosure; Fig 8; 94pp: English.  
XX  
XX The present sequence is the translated sequence found at the cloning  
CC junction in the pET30a vector of nematode Haemonchus contortus  
CC clone 65e. This clone encodes a novel protein. This organism is a  
CC parasite found in the stomach of its host (sheep and goats in  
CC particular). The protein can be used in a vaccine against other  
CC helminths, as well as Haemonchus contortus, including trematodes,  
CC cestodes, nematodes and acanthocephala. These are all capable of causing  
CC severe illness in their hosts, which include sheep, pigs, goats, cattle,  
CC horses, donkeys, dogs, cats, guinea pigs and cage-birds, along with

CC humans. Antibodies to the protein can be used to diagnose infection.  
CC Note: This sequence is stated as being the same as that shown in SEQ ID  
CC NO: 6 of the specification (see AAB26109). However, this sequence is  
CC shorter than the one shown here.  
XX  
SQ Sequence 76 AA;  
Query Match 11.6%; Score 50; DB 21; Length 76;  
Best Local Similarity 100.0%; Pred. No. 3e-42;  
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MHHHHHSSGLVPRGSGMKETAATAKFERQHMDSPDLGTDGDDKAMADIGS 50  
DB 1 MHHHHHSSGLVPRGSGMKETAATAKFERQHMDSPDLGTDGDDKAMADIGS 50  
RESULT 8  
AAB85253  
ID AAB85253 standard; Protein; 167 AA.  
XX  
AC AAB85253;  
XX  
XX 07-SEP-2001 (first entry)  
XX  
XX Thioredoxin functional fragment.  
DE  
XX Thioredoxin; porphobilinogen synthase; T-PPS; enzyme; herbicide;  
XX delta-aminolevulinic acid; plant growth inhibitor.  
XX  
XX Lycopersicon esculentum.  
XX  
XX WO200146446-A1.  
XX  
XX 28-JUN-2001.  
XX  
XX 19-DEC-2000; 2000WO-US34584.  
XX  
XX 22-DEC-1999; 99US-0171785.  
XX  
XX (PARA-) PARADIGM GENETICS INC.  
XX  
XX Crawford JM, Rice J, Sevala V, Stewart S;  
PI  
XX WPI: 2001-418081/44.  
DR N-PSDB; AAH22802.  
XX  
XX Novel plant thioredoxin-porphobilinogen synthase or porphobilinogen  
PT synthase polypeptides, useful for identifying compounds for use as  
PT herbicides by inhibiting enzymatic activity of the polypeptides -  
XX  
XX Example 1; Page 19; 25pp: English.  
XX  
XX The invention provides novel DNA sequences encoding enzymes such as plant  
CC thioredoxin-porphobilinogen synthase (T-PPS) and plant porphobilinogen  
CC synthase (PPS). Methods of determining the enzymatic activity of T-PPS or  
CC PPS or its functional fragment are provided that involve contacting  
CC delta-aminolevulinic acid with the protein or its functional fragment and  
CC measuring the amount of porphobilinogen formed from it. Compounds which  
CC can modify the enzymatic activity of T-PPS or PPS can also be identified  
CC similarly, which are useful for inhibiting plant growth by inhibiting  
CC enzymatic activity of T-PPS or PPS or its functional fragment. The  
CC compounds thus identified are useful as herbicides. The present sequence  
CC represents the thioredoxin functional fragment.  
XX  
SQ Sequence 167 AA;  
Query Match 11.6%; Score 50; DB 22; Length 167;  
Best Local Similarity 100.0%; Pred. No. 6.3e-42;  
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MHHHHHSSGLVPRGSGMKETAATAKFERQHMDSPDLGTDGDDKAMADIGS 50  
DB 116 MHHHHHSSGLVPRGSGMKETAATAKFERQHMDSPDLGTDGDDKAMADIGS 165

RESULT 9  
 AAB85254  
 ID AAB85254 standard; Protein; 167 AA.  
 XX AC AAB85254;  
 XX DT 07-SEP-2001 (first entry)  
 XX DE Thioresoxin functional fragment.  
 XX KW Thioresoxin; porphobilinogen synthase; T-PPS; PPS; enzyme; herbicide;  
 XX KW delta-aminolevulinic acid; plant growth inhibitor.  
 XX OS Lycopersicon esculentum.  
 XX PN WO200146446-A1.  
 XX PD 28-JUN-2001.  
 XX PF 19-DEC-2000; 2000WO-US34584.  
 XX PR 22-DEC-1999; 99US-0171785.  
 XX PA (PARA-) PARADIGM GENETICS INC.  
 XX PI Crawford JM, Rice J, Sevala V, Stewart S;  
 XX DR WPI; 2001-418081/44.  
 XX DR N-PSDB; AAB22802.  
 XX KW Novel plant thioresoxin-porphobilinogen synthase or porphobilinogen  
 PT synthase polypeptides, useful for identifying compounds for use as  
 PT herbicides by inhibiting enzymatic activity of the polypeptides -  
 XX Example 1; Page 20; 25pp; English.  
 XX CC The invention provides novel DNA sequences encoding enzymes such as plant  
 CC thioresoxin-porphobilinogen synthase (T-PPS) and plant porphobilinogen  
 CC synthase (PPS). Methods of determining the enzymatic activity of T-PPS or  
 CC PPS or its functional fragment are provided that involves contacting  
 CC delta-aminolevulinic acid with the protein or its functional fragment and  
 CC measuring the amount of porphobilinogen formed from it. Compounds which  
 CC can modify the enzymatic activity T-PPS or PPS can also be identified  
 CC similarly, which are useful for inhibiting plant growth by inhibiting  
 CC enzymatic activity of T-PPS or PPS or its functional fragment. The  
 CC compounds thus identified are useful as herbicides. The present sequence  
 CC represents the thioresoxin functional fragment.  
 XX SQ Sequence 167 AA;  
 Query Match 11.6%; Score 50; DB 22; Length 167;  
 Best Local Similarity 100.0%; Pred. No. 6.3e-42;  
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 MHHHHHSSGLVPRGSGMKETAARQHMDSPLGTDDDDKAMADIGS 50  
 Db 116 MHHHHHSSGLVPRGSGMKETAARQHMDSPLGTDDDDKAMADIGS 165  
 RESULT 10  
 AAE13033  
 ID AAE13033 standard; Protein; 289 AA.  
 XX AC AAE13033;  
 XX DT 28-JAN-2002 (first entry)  
 XX KW Helicobacter pylori, His-HOP38(-11) polypeptide.  
 DE Polypeptide delivery system; immune stimulating complex; ISCOM;  
 KW cholesterol; saponin; phospholipid; medicament; vaccine; therapy;  
 XX

KW Helicobacter pylori infection; HOP38; anti-bacterial.  
 OS Helicobacter pylori.  
 XX WO200176625-A1.  
 XX PD 18-OCT-2001.  
 XX PF 09-APR-2001; 2001WO-SF00799.  
 XX PR 12-APR-2000; 2000GB-0008879.  
 XX PA (ASTR ) ASTRAZENECA AB.  
 XX PI Shapiro A, Sanyal G;  
 XX DR WPI; 2001-663016/76.  
 XX DR N-PSDB; AAD21371.  
 XX KW Producing a polypeptide delivery system useful in a vaccine to treat  
 PT infection by mixing together the polypeptide, cholesterol, saponin, and  
 PT a phospholipid in presence of a nonionic detergent and a second  
 PT detergent -  
 XX Disclosure; Page 40; 43pp; English.  
 XX CC The invention relates to producing a polypeptide delivery system  
 CC comprising an immune stimulating complex (ISCOM) coupled to a  
 CC polypeptide. The method involves mixing the polypeptide, cholesterol,  
 CC saponin and a phospholipid in the presence of a nonionic detergent  
 CC and a second detergent to form a solution and removing the detergent  
 CC from the mixture to form the ISCOM. The polypeptide delivery system  
 CC is used as a medicament for the manufacture of a vaccine for  
 CC administration to mammalian patients, to treat and prevent  
 CC Helicobacter pylori infection. The present sequence is His-HOP38(-11),  
 CC a Helicobacter pylori polypeptide used in the polypeptide delivery  
 CC system.  
 XX SQ Sequence 289 AA;  
 Query Match 11.6%; Score 50; DB 22; Length 289;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-41;  
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 MHHHHHSSGLVPRGSGMKETAARQHMDSPLGTDDDDKAMADIGS 50  
 Db 1 MHHHHHSSGLVPRGSGMKETAARQHMDSPLGTDDDDKAMADIGS 50  
 RESULT 11  
 AAB47732  
 ID AAB47732 standard; Protein; 289 AA.  
 XX AC AAB47732;  
 XX DT 30-JAN-2002 (first entry)  
 XX DE HOP38 #5.  
 XX KW H. pylori; HOP38; polypeptide delivery system; polar head group;  
 KW immune stimulating complex; ISCOM; antigenic fragment; saponin;  
 KW cholesterol; phospholipid; detergent.  
 XX OS Helicobacter pylori.  
 XX PN WO200176623-A1.  
 XX PD 18-OCT-2001.  
 XX PF 09-APR-2001; 2001WO-SF00800.  
 XX PR 12-APR-2000; 2000GB-0008877.  
 XX

PA (ASTR ) ASTRAZENECA AB.

PI Shaprio A, Sanyal G;

DR WPI; 2002-025884/03.

DR N-PSDB; AAH43782.

XX Production of a polypeptide delivery system useful as a medicament comprises mixing together the polypeptide, cholesterol, saponin, and a phospholipid with a polar head group, in the presence of a detergent.

PS Disclosure; Page 45; 48pp; English.

XX The sequences given in ABA47728-32 are H. pylori HOP38 proteins which were used in the method of the invention. The method of the invention is a process for production of a polypeptide delivery system comprising an immune stimulating complex (ISCOM) coupled to a polypeptide of H. pylori or its antigenic fragment. The method comprises: mixing the polypeptide, cholesterol, a saponin, and a phospholipid having a polar head group; and removing the detergent from the mixture to form an ISCOM. The method has a broad applicability to polypeptides, including polypeptides that are unsuitable to prior art processes.

SQ Sequence 289 AA;

Query Match 11.6%; Score 50; DB 23; Length 289;

Best Local Similarity 100.0%; Pred. No. 1.1e-41;

Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHSSGLVPRGSGMKETAARFERQHMDSPDLGTDGDDKAMADIGS 50

|||||

DB 1 MHHHHHSSGLVPRGSGMKETAARFERQHMDSPDLGTDGDDKAMADIGS 50

RESULT 12

ABG67266

ID ABG67266 standard; Protein; 402 AA.

XX

AC ABG67266;

XX

DT 24-SEP-2002 (first entry)

XX

DE Polythoa 2 fluorescent protein encoded by plasmid pGR3.

XX

KW Coral; Anthozoa species; fluorescence emission; fluorescence absorbance;

KW fluorescence resonance energy transfer; FRET; fluorescent labelling;

KW Polythoa 2 fluorescent protein; mutant; mutein.

XX

OS Polythoa sp.

OS Synthetic.

XX

PN WO200242323-A2.

XX

PD 30-MAY-2002.

XX

PF 22-NOV-2001; 2001WO-EPI3604.

XX

PR 22-NOV-2000; 2000GB-0028495.

XX

PR 22-NOV-2000; 2000US-252790P.

XX

PA (DEVG-) DEVGEN NV.

XX

PI Nys G, Plaetinck G, Bogaert T;

XX

DR WPI; 2002-500276/53.

XX

DR N-PSDB; ABK96468.

XX

PT Novel isolated fluorescent protein for producing fluorescence resonance

PT energy transfer for use in vivo labelling studies, is capable of

PT emitting fluorescence upon irradiation by incident light

XX

PS Claim 7; Fig 4; 96pp; English.

XX

CC The present invention relates to novel coral fluorescent proteins

CC isolated from two brightly fluorescent Anthozoa species (Polythoa

CC and Discosoma species), and the polynucleotide sequences encoding

CC them. The fluorescent proteins of the invention are capable of

CC emitting fluorescence upon incident light irradiation, where the

CC maximal absorbance of the incident light is 440-480 nm, and maximal

CC fluorescence emission is 470-510 nm. The fluorescent proteins are

CC useful for producing fluorescence resonance energy transfer (FRET),

CC between a donor and acceptor fluorescent protein molecule. The

CC fluorescent proteins are useful for in vivo labelling studies, as

CC a label and/or marker, in particular as a genetic marker and/or

CC expression marker, in the fields of (micro-) biology, biochemistry

CC and/or molecular biology. They are also useful for monitoring

CC expression of proteins within biological systems and subcellular

CC localisation or trafficking of proteins. The polynucleotide

CC sequences encoding the fluorescent proteins are useful for in vitro

CC applications such as hybridisation and/or immunological assays. They

CC are also useful for labelling polypeptides of interest. The present

CC sequence represents a Polythoa 2 fluorescent protein encoded by a

CC plasmid as described in the examples of the present invention.

XX

SQ Sequence 402 AA;

Query Match 11.6%; Score 50; DB 23; Length 402;

Best Local Similarity 100.0%; Pred. No. 1.4e-41;

Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHSSGLVPRGSGMKETAARFERQHMDSPDLGTDGDDKAMADIGS 50

|||||

DB 116 MHHHHHSSGLVPRGSGMKETAARFERQHMDSPDLGTDGDDKAMADIGS 165

|||||

RESULT 13

ABG67267

ID ABG67267 standard; Protein; 407 AA.

XX

AC ABG67267;

XX

DT 24-SEP-2002 (first entry)

XX

DE Polythoa 2 fluorescent protein encoded by plasmid pGR7.

XX

KW Coral; Anthozoa species; fluorescence emission; fluorescence absorbance;

KW fluorescence resonance energy transfer; FRET; fluorescent labelling;

KW Polythoa 2 fluorescent protein; mutant; mutein.

XX

OS Polythoa sp.

OS Synthetic.

XX

PN WO200242323-A2.

XX

PD 30-MAY-2002.

XX

PF 22-NOV-2001; 2001WO-EPI3604.

XX

PR 22-NOV-2000; 2000GB-0028495.

XX

PR 22-NOV-2000; 2000US-252790P.

XX

PA (DEVG-) DEVGEN NV.

XX

PI Nys G, Plaetinck G, Bogaert T;

XX

DR WPI; 2002-500276/53.

XX

DR N-PSDB; ABK96472.

XX

PT Novel isolated fluorescent protein for producing fluorescence resonance

PT energy transfer for use in vivo labelling studies, is capable of

PT emitting fluorescence upon irradiation by incident light

XX

PS Claim 7; Fig 5; 96pp; English.

XX

XX

CC The present invention relates to novel coral fluorescent proteins

CC isolated from two brightly fluorescent Anthozoa species (Polythoa



QY 1 MHHHHHSSGLVPRSGMKETAATAKFERQHMDSPDLGTDGDDKAMADIGS 50  
|||||  
Db 116 MHHHHHSSGLVPRSGMKETAATAKFERQHMDSPDLGTDGDDKAMADIGS 165

## RESULT 16

AAAY79067  
ID AAY79067 standard; Protein; 467 AA.

XX  
AC AAY79067;

XX 20-JUN-2000 (first entry)

DE Murine secreted protein acidic and rich in cysteine amino acid sequence.

KW Secreted protein acidic and rich in cysteine; SPARC; mouse; nerve cell;  
KW neurocyte cell adhesion; cell protrusion retraction; cell shrinkage;  
KW nervous system disease; epilepsy; arteriosclerosis; wound healing.

XX Mus sp.

XX JP3012930-B1.

XX 28-FEB-2000.

XX 26-FEB-1999; 99JP-0049708.

XX 26-FEB-1999; 99JP-0049708.

XX (AGEN ) KOGYO GIUTSUINCHO.

XX WPI; 2000-306484/27.

XX N-PSDB; AAZ98759.

XX Drug composition for suppressing neurocyte cell adhesion, generating  
PT cell migration and promoting shrinkage retraction of the cell contains  
PT specific amino acid sequence -

XX Claim 1; Fig 4; 21pp; Japanese.

XX This sequence represents the amino acid sequence of the murine protein  
CC referred to as SPARC (secreted protein acidic and rich in cysteine). The  
CC protein is used as the active ingredient in a drug composition for  
CC suppressing neurocyte cell adhesion, generating cell migration and  
CC promoting a nerve protrusion shrinkage reaction. The composition can be  
CC used in wound healing and also to treat diseases of the nervous system,  
CC arteriosclerosis, and epilepsy.

SQ Sequence 467 AA;

Query Match 11.6%; Score 50; DB 21; Length 467;  
Best Local Similarity 100.0%; Pred. No. 1.7e-41;  
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHSSGLVPRSGMKETAATAKFERQHMDSPDLGTDGDDKAMADIGS 50  
|||||  
Db 116 MHHHHHSSGLVPRSGMKETAATAKFERQHMDSPDLGTDGDDKAMADIGS 165

## RESULT 17

AAAB85251  
ID AAB85251 standard; Protein; 551 AA.

XX  
AC AAB85251;

XX 07-SEP-2001 (first entry)

XX Plant thioresoxin-porphobilinogen synthase (T-PPS).

XX Thioresoxin; porphobilinogen synthase; T-PPS; pps; enzyme; herbicide;  
KW delta-aminolevulinic acid; plant growth inhibitor.

XX Lycopersicon esculentum.

XX Key Location/Qualifiers  
FH Protein 1..167  
FT /note= "thioresoxin functional fragment"  
FT 168..551  
FT /note= "pps"

XX WO200146446-A1.

XX 28-JUN-2001.

XX 19-DEC-2000; 2000WO-US34584.

XX 22-DEC-1999; 99US-0171785.

XX (PARA-) PARADIGM GENETICS INC.

XX Crawford JM, Rice J, Sevala V, Stewart S;

XX WPI; 2001-418081/44.

XX N-PSDB; AAH22901.

XX Novel plant thioresoxin-porphobilinogen synthase or porphobilinogen  
PT synthase polypeptides, useful for identifying compounds for use as  
PT herbicides by inhibiting enzymatic activity of the polypeptides -

XX Disclosure; Page 16-17; 25pp; English.

XX The invention provides novel DNA sequences encoding enzymes such as plant  
CC thioresoxin-porphobilinogen synthase (T-PPS) and plant porphobilinogen  
CC synthase (PPS). Methods of determining the enzymatic activity of T-PPS or  
CC PPS or its functional fragment are provided that involves contacting  
CC delta-aminolevulinic acid with the protein or its functional fragment and  
CC measuring the amount of porphobilinogen formed from it. Compounds which  
CC can modify the enzymatic activity T-PPS or PPS can also be identified  
CC similarly, which are useful for inhibiting plant growth by inhibiting  
CC enzymatic activity of T-PPS or PPS or its functional fragment. The  
CC compounds thus identified are useful as herbicides. The present sequence  
CC represents the fusion protein T-PPS.

SQ Sequence 551 AA;

Query Match 11.6%; Score 50; DB 22; Length 551;  
Best Local Similarity 100.0%; Pred. No. 1.9e-41;  
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHSSGLVPRSGMKETAATAKFERQHMDSPDLGTDGDDKAMADIGS 50  
|||||  
Db 116 MHHHHHSSGLVPRSGMKETAATAKFERQHMDSPDLGTDGDDKAMADIGS 165

## RESULT 18

AAAB85252  
ID AAB85252 standard; Protein; 551 AA.

XX  
AC AAB85252;

XX 07-SEP-2001 (first entry)

XX Plant thioresoxin-porphobilinogen synthase (T-PPS).

XX Thioresoxin; porphobilinogen synthase; T-PPS; pps; enzyme; herbicide;  
KW delta-aminolevulinic acid; plant growth inhibitor.

XX Lycopersicon esculentum.

XX Key Location/Qualifiers

FH Protein 1..167

FT /note= "thioresoxin functional fragment"

FT 168..551

FT /note= "pps"

XX WO200146446-A1.

```

XX PD 28-JUN-2001.
XX PF
XX PF 19-DEC-2000; 2000WO-US34584.
XX PR 22-DEC-1999; 99US-0171785.
XX PA (PARA-) PARADIGM GENETICS INC.
XX PI Crawford JM, Rice J, Sevala V, Stewart S;
XX PI WPI: 2001-418081/44.
XX DR N-PSDB; AAH22801.
XX PT Novel plant thioredoxin-porphobilinogen synthase or porphobilinogen
XX PT synthase polypeptides, useful for identifying compounds for use as
XX PT herbicides by inhibiting enzymatic activity of the polypeptides -
XX PS Claim 4; Page 17-18; 25pp; English.
XX CC The invention provides novel DNA sequences encoding enzymes such as plant
XX CC thioredoxin-porphobilinogen synthase (T-PPS) and plant porphobilinogen
XX CC synthase (PPS). Methods of determining the enzymatic activity of T-PPS or
XX CC PPS or its functional fragment are provided that involves contacting
XX CC delta-aminolevulinic acid with the protein or its functional fragment and
XX CC measuring the amount of porphobilinogen formed from it. Compounds which
XX CC can modify the enzymatic activity T-PPS or PPS can also be identified
XX CC similarly, which are useful for inhibiting plant growth by inhibiting
XX CC enzymatic activity of T-PPS or PPS or its functional fragment. The
XX CC compounds thus identified are useful as herbicides. The present sequence
XX CC represents the fusion protein T-PPS.
XX SQ Sequence 551 AA;
Query Match 11.6%; Score 50; DB 22; Length 551;
Best Local Similarity 100.0%; Pred. No. 1.9e-41;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MHHHHHSSGLVPRGSGMKETAFAKFERQHMDSPDLGTDGDDDKAMADIGS 50
Db 116 MHHHHHSSGLVPRGSGMKETAFAKFERQHMDSPDLGTDGDDDKAMADIGS 165
RESULT 19
AA11532
ID AA11532 standard; Protein: 692 AA.
XX AC AAB11532;
XX DT 12-OCT-2000 (first entry)
XX DE SEN virus protein fragment SEQ ID NO: 104.
XX KW SEN virus; SENV; gastrointestinal tract disorder; inflammatory disease;
XX KW proliferative disorder; hepatopathy; hepatitis; viral infection;
XX KW vaccination; gene therapy.
XX OS Hepatitis virus.
XX PN WO200028039-A2.
XX PD 18-MAY-2000.
XX PF 09-NOV-1999; 99WO-EF08566.
XX PR 10-NOV-1998; 98IT-MI02437.
XX PR 30-APR-1999; 99IT-MI00923.
XX PR 14-MAY-1999; 99EP-0830298.
XX PR 16-JUL-1999; 99EP-0113932.
XX PA (DIAS-) DIASORIN SRL.
XX PI Primi D, Fiordalisi G, Mantero GL, Mattioli S, Sottini A;

```

```

PI Bonelli F, Vaglini L, Olivero P, Dal Corso A, Bonelli M;
XX DR WPI: 2000-376551/32.
XX PT Nucleic acids representing the genome of the SEN virus (SENV) and
XX PT encoded proteins, useful for treatment of hepatopathies, inflammatory
XX PT diseases and proliferative disorders such as cancer -
XX PS Example 16; Page 78; 392pp; English.
XX CC The present invention is concerned with the sequence of the genome of the
XX CC SEN virus (SENV), and the proteins encoded by it. SENV is thought to be
XX CC the cause of hepatopathies which are not linked to the presence of the
XX CC hepatitis A, B and E viruses in man. The genome and proteins of this
XX CC virus can be used in gene therapy and vaccination against the virus,
XX CC which also causes disorders of the gastrointestinal tract, including
XX CC Crohn's disease and lupus erythematosus, inflammatory diseases, and
XX CC proliferative disorders such as cancer.
XX SQ Sequence 692 AA;
Query Match 11.6%; Score 50; DB 21; Length 692;
Best Local Similarity 100.0%; Pred. No. 2.4e-41;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MHHHHHSSGLVPRGSGMKETAFAKFERQHMDSPDLGTDGDDDKAMADIGS 50
Db 1 MHHHHHSSGLVPRGSGMKETAFAKFERQHMDSPDLGTDGDDDKAMADIGS 50
RESULT 20
AA48247
ID AA48247 standard; Protein: 824 AA.
XX AC AAM48247;
XX DT 21-MAR-2002 (first entry)
XX DE Thioredoxin/deoxyxylulose 5-phosphate synthase fusion protein.
XX KW DXPS: deoxyxylulose 5-phosphate synthase; pyruvate; thioredoxin;
XX KW glyceraldehyde 3-phosphate; plant growth modulator;
XX KW microbial growth modulator; enzyme.
XX OS Chimeric - Arabidopsis thaliana.
XX OS Chimeric - Synthetic.
XX PH Key Location/Qualifiers
XX FT Region 1..165
XX FT /note= "Thioredoxin protein"
XX FT Region 166..824
XX FT /note= "Truncated DXPS protein"
XX PN US6326164-B1.
XX PD 04-DEC-2001.
XX PF 27-JUL-2000; 2000US-0626589.
XX PR 27-JUL-2000; 2000US-0626589.
XX PA (PARA-) PARADIGM GENETICS INC.
XX PI Rice JW, Kloti AS, Crawford JM, Lanning B, Stewart SJ;
XX DR WPI: 2002-121106/16.
XX DR N-PSDB; ABA95618.
XX PT Assay for determining deoxyxylulose 5-phosphate synthase activity by
XX PT measuring pyruvate depletion, useful for screening compounds that
XX PT inhibit or enhance this activity which is useful for modulating plant
XX PT and microbial growth -
XX

```



PS Claim 1; Columns 19-24; 26pp; English.

XX The present sequence is a fusion protein comprising a truncated

CC deoxyxylulose 5-phosphate synthase (DXPS) from Arabidopsis thaliana and

CC thioredoxin (trxA). The N-terminal 58 amino acids of the DXPS protein

CC were removed to generate the truncated protein. The truncated DXPS protein

CC possesses DXPS activity. The full-length DXPS is given in AAM48245. The

CC present invention relates to a method for determining DXPS activity. The

CC method comprises contacting pyruvate, and optionally glyceraldehyde

CC 3-phosphate with DXPS and then determining the concentration of remaining

CC pyruvate and/or glyceraldehyde 3-phosphate. The method is useful for

CC screening for inhibitors and enhancers of DXPS activity which will have

CC use as modulators of plant and microbial growth.

XX SQ Sequence 824 AA;

Query Match 11.6%; Score 50; DB 23; Length 824;

Best Local Similarity 100.0%; Pred. No. 2.8e-41;

Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 MHHHHHSSGLVPRGSGMKETAARFQRHMDSPDLGTDGDDDDKAMADIGS 50

Db 116 MHHHHHSSGLVPRGSGMKETAARFQRHMDSPDLGTDGDDDDKAMADIGS 165

RESULT 21

AAR41355

ID AAR41355 standard; peptide; 49 AA.

XX

AC AAR41355;

XX

DT 04-MAR-1994 (first entry)

XX

DE MSP1EGF1B EGF1-like domain variant.

XX

KW Epidermal growth factor 1; merozoite surface protein 1; malaria;

KW vaccine.

XX

OS Plasmodium yoelii.

XX Key Location/Qualifiers

FH Cleavage-site 1

FT /note= "introduced to facilitate cleavage

FT from recombinant protein"

XX PN WO9317107-A.

XX

XX 02-SEP-1993.

XX

XX 22-FEB-1993; 93WO-GB00367.

XX

XX 22-FEB-1992; 92GB-0003821.

XX

XX (MEDI-) MEDICAL RES COUNCIL.

XX

PI Blackman MJ, Chappel JA, Holder AA;

XX

DR WPI; 1993-288413/36.

XX

PT Allelic variants of epidermal growth factor 1- or 2-like domains - of

PT merozoite surface protein 1, produced recombinantly for malaria

PT vaccines

XX

PS Claim 1; Fig 1b; 35pp; English.

XX The sequence is that of an allelic variant of a merozoite surface

CC protein-1 epidermal growth factor (EGF) 1 like domain. It may be

CC used alone or as part of a fusion protein of EGF-1-like and

CC EGF-2-like domains in vaccines against malaria. When expressed

CC recombinantly it is produced in a form indistinguishable from

CC that in the native protein.

XX

XX SQ Sequence 49 AA;

Query Match 11.1%; Score 48; DB 14; Length 49;

Best Local Similarity 100.0%; Pred. No. 2e-40;

Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 338 NISQHCQVKKQCPENSGCFRHLDERECKCLLNKQEGDKCVENPNPT 385

Db 2 NISQHCQVKKQCPENSGCFRHLDERECKCLLNKQEGDKCVENPNPT 49

RESULT 22

AAB37608

ID AAB37608 standard; protein; 96 AA.

XX

AC AAB37608;

XX

DT 27-FEB-2001 (first entry)

XX

DE Merozoite surface protein-1.

XX

KW Merozoite surface protein; protozoacide; vaccine; malaria.

XX

OS Plasmodium falciparum.

XX WO200063245-A2.

XX 26-OCT-2000.

XX 20-APR-2000; 2000WO-GB01558.

XX 20-APR-1999; 99GB-0009072.

PR 13-MAY-1999; 99US-0311817.

PR 25-MAY-1999; 99CA-2271451.

XX (MEDI-) MEDICAL RES COUNCIL.

XX PI Holder A, Birdsall B, Feeney J, Morgan W, Syed S, Uthaiipibull C;

XX WPI; 2001-015762/02.

XX Novel variants of the C-terminal fragment of Plasmodium merozoite

PT surface protein-1, useful as vaccines for treating or preventing

PT malaria -

XX Example 2; Page 48; 126pp; English.

XX The present invention relates to non-natural variants of a C-terminal

CC fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The

CC non-natural variants have reduced affinity for at least 1 antibody

CC capable of blocking a second antibody that inhibits the proteolytic

CC cleavage of Plasmodium MSP-1\_4\_2, and has the same affinity for at least

CC one third antibody that inhibits the proteolytic cleavage of Plasmodium

CC MSP-1\_4\_2, compared to natural MSP-1\_1\_9. The present sequence is the

CC wild-type MSP-1 protein. This sequence was used to generate the variants

CC of the present invention. The non-natural variants of the present

CC invention are useful for immunising a mammal against malaria, and can be

CC used to treat malaria.

XX SQ Sequence 96 AA;

Query Match 10.7%; Score 46; DB 22; Length 96;

Best Local Similarity 100.0%; Pred. No. 4e-38;

Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 352 NSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENNGCGDADA 397

Db 15 NSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENNGCGDADA 60

RESULT 23

AAB37609

ID AAB37609 standard; Protein; 108 AA.

XX



XX PD 06-FEB-2002.  
XX PF 02-AUG-2000; 2000EP-0202750.  
XX PR 02-AUG-2000; 2000EP-0202750.  
XX PA (UYRO-) UNIV ROTTERDAM ERASMUS.  
XX PI Scholte BJ;  
XX DR WPI: 2002-270869/32.  
XX DR N-PSDB; AAD32715.  
XX PT New gene delivery vehicle comprising a nucleic acid binding domain and  
XX PT an integrin binding domain, useful for delivering therapeutic molecule  
XX PT to a cell exposing integrin and enhance efficacy of gene delivery using  
XX PT integrin receptors  
XX PS Disclosure; Fig 1A; 32pp; English.  
XX CC The invention relates to a gene delivery vehicle comprising a general  
XX CC nucleic acid binding domain and an integrin binding domain. The  
XX CC invention also provides modular protein (SPKR)4inv, designed to mediate  
XX CC integrin targeted gene delivery. The gene delivery vehicle is useful  
XX CC for delivering therapeutic molecule to a cell exposing integrin, and  
XX CC subsequently enhance the efficacy of gene delivery using integrin  
XX CC receptors. The present sequence is recombinant (SPKR)4inv protein.  
XX SQ Sequence 284 AA;  
Query Match 10.7%; Score 46; DB 23; Length 284;  
Best Local Similarity 100.0%; Pred. No. 1.1e-37;  
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MHHHHHSSGLVPRGSGMKETAARKFERQHMDSPDLGTDGDDKAMA 46  
|||||  
DB 1 MHHHHHSSGLVPRGSGMKETAARKFERQHMDSPDLGTDGDDKAMA 46  
RESULT 26  
AAY09372  
ID AAY09372 standard; Protein; 355 AA.  
XX AC AAY09372;  
XX DT 31-AUG-1999 (first entry)  
XX DE Merozoite surface protein MSP-1-42.  
XX KW MSP-1; merozoite surface protein; malaria; vaccine;  
XX KW protein engineering; protein expression; codon usage;  
XX KW transgenic animal.  
XX OS Plasmodium falciparum.  
XX PN WO9920774-A2.  
XX PD 29-APR-1999.  
XX PF 20-OCT-1998; 98WO-US22226.  
XX PR 15-MAY-1998; 98US-0085649.  
XX PR 20-OCT-1997; 97US-0062592.  
XX PA (GENZ ) GENZYME TRANSGENICS CORP.  
XX PI Chen LH, Meade H;  
XX PI WPI: 1999-288313/24.  
XX DR P-PSDB; AAX56008.  
XX PT Modified malarial protein for use in anti-malarial vaccines

XX PS Example; Fig 1; 35pp; English.  
XX CC The present sequence represents a 42 kDa C-terminal portion of  
XX CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an  
XX CC important target for the development of a vaccine against  
XX CC Plasmodium falciparum. The nucleic acid sequence encoding MSP-1-42  
XX CC has been modified (see AAX56008) compared to the native sequence (see  
XX CC AAX56009) such that 306 nucleotide positions have been replaced to  
XX CC lower the AT content (from 76 to 49.7%) and to eliminate 10 mRNA  
XX CC instability motifs while maintaining the same protein amino acid  
XX CC sequence. These alterations allow MSP-1-42 to be expressed in  
XX CC mammalian cell culture and in transgenic mice. Native MSP-1-12  
XX CC is known to be difficult to express in cell culture systems,  
XX CC mammalian cell culture systems or in transgenic animals. The  
XX CC invention allows expression of MSP-1 protein in the milk of  
XX CC transgenic animals, and also provides a DNA vaccine comprising a  
XX CC vector containing the altered MSP-1-42 sequence.  
XX SQ Sequence 355 AA;  
Query Match 10.7%; Score 46; DB 20; Length 355;  
Best Local Similarity 100.0%; Pred. No. 1.4e-37;  
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 352 NSGCFRHLDERECKLLNYKQEGDKCVENPNPTCENNGGCDADA 397  
|||||  
DB 277 NSGCFRHLDERECKLLNYKQEGDKCVENPNPTCENNGGCDADA 322  
RESULT 27  
AAY05832  
ID AAY05832 standard; Protein; 355 AA.  
XX AC AAY05832;  
XX DT 02-AUG-1999 (first entry)  
XX DE Merozoite surface protein MSP-1-42.  
XX KW MSP-1; merozoite surface protein; malaria; vaccine;  
XX KW protein engineering; protein expression; codon usage;  
XX KW transgenic animal.  
XX OS Plasmodium falciparum.  
XX PN WO9920766-A2.  
XX PD 29-APR-1999.  
XX PF 20-OCT-1998; 98WO-US22225.  
XX PR 15-MAY-1998; 98US-0085649.  
XX PR 20-OCT-1997; 97US-0062592.  
XX PA (GENZ ) GENZYME TRANSGENICS CORP.  
XX PI Chen LH, Meade H;  
XX PI WPI: 1999-302742/25.  
XX DR N-PSDB; AAX25586.  
XX PT New modified recombinant nucleic acid sequences useful for producing  
XX PT malarial DNA vaccine  
XX PS Disclosure; Fig 1; 43pp; English.  
XX CC The present sequence represents a 42 kDa C-terminal portion of  
XX CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an  
XX CC important target for the development of a vaccine against  
XX CC Plasmodium falciparum. The nucleic acid sequence encoding MSP-1-42  
XX CC has been modified (see AAX25586) compared to the native sequence (see  
XX CC AAX25587) such that 306 nucleotide positions have been replaced to

CC lower the AT content (from 76 to 49.7%) and to eliminate 10 mRNA  
 CC instability motifs while maintaining the same protein amino acid  
 CC sequence. These alterations allow MSP-1-42 to be expressed in  
 CC mammalian cell culture and in transgenic mice. The invention  
 CC provides modified recombinant nucleic acid sequences and methods for  
 CC increasing the mRNA levels and protein expression of proteins that  
 CC are difficult to express in cell culture systems, mammalian cell  
 CC culture systems or in transgenic animals. The preferred difficult  
 CC protein candidates for expression are those derived from lower  
 CC organisms such as parasites, bacteria and viruses that have DNA  
 CC coding sequences of high AT content or which have mRNA instability  
 CC motifs or rare codons relative to the recombinant expression system  
 CC to be used. The invention allows expression of MSP-1 protein in  
 CC the milk of transgenic animals, and also provides a DNA vaccine  
 CC comprising a vector containing the altered MSP-1-42 sequence.

XX Sequence 355 AA;

Query Match 10.7%; Score 46; DB 20; Length 355;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-37;  
 Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 NSGCFRHLDERECKLLNYKQEGDKCVENPPTCENNNGCCDADA 397  
 |||||||||||||||||||||||||||||||||||||||||||||||||

Db 277 NSGCFRHLDERECKLLNYKQEGDKCVENPPTCENNNGCCDADA 322

RESULT 28

AAV09373

ID AAY09373 standard; Protein; 361 AA.

XX AC AAY09373;

XX DT 31-AUG-1999 (first entry)

XX DE Merozoite surface protein MSP-1-42.

XX KW MSP-1; merozoite surface protein; malaria; vaccine;  
 KW protein engineering; protein expression; codon usage;  
 KW transgenic animal.

XX OS Plasmodium falciparum.

XX PN WO9920774-A2.

XX PD 29-APR-1999.

XX PF 20-OCT-1998; 98WO-US22226.

XX PR 15-MAY-1998; 98US-0085649.

XX PR 20-OCT-1997; 97US-0062592.

XX PA (GENZ ) GENZYME TRANSGENICS CORP.

XX PI Chen LH, Meade H;

XX DR WPI; 1999-288313/24.

XX DR P-PSDB; AAX56009.

XX PT Modified malarial protein for use in anti-malarial vaccines

XX PS Example; Fig 2; 35pp; English.

XX CC This present sequence comprises a 42 kDa C-terminal portion of  
 CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an  
 CC important target for the development of a vaccine against  
 CC Plasmodium falciparum. The C-terminal end of the sequence is  
 CC modified to include a 6xHis tag. A nucleic acid (see AAX56009)

CC encoding MSP-1-42 has been modified according to a method  
 CC of the invention in order to improve expression in mammalian cells  
 CC and in transgenic animals by reducing the AT content and removing  
 CC mRNA instability motifs. The invention allows expression of  
 CC MSP-1-42 in the milk of transgenic animals, and also provides a DNA

CC vaccine comprising a vector containing the altered MSP-1-42 nucleic  
 CC acid.  
 XX Sequence 361 AA;

Query Match 10.7%; Score 46; DB 20; Length 361;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-37;  
 Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 NSGCFRHLDERECKLLNYKQEGDKCVENPPTCENNNGCCDADA 397

Db 277 NSGCFRHLDERECKLLNYKQEGDKCVENPPTCENNNGCCDADA 322

RESULT 29

AAV05833

ID AAY05833 standard; Protein; 361 AA.

XX AC AAY05833;

XX DT 02-AUG-1999 (first entry)

XX DE Merozoite surface protein MSP-1-42.

XX KW MSP-1; merozoite surface protein; malaria; vaccine;  
 KW protein engineering; protein expression; codon usage;  
 KW transgenic animal.

XX OS Plasmodium falciparum.

XX PN WO9920766-A2.

XX PD 29-APR-1999.

XX PF 20-OCT-1998; 98WO-US22225.

XX PR 15-MAY-1998; 98US-0085649.

XX PR 20-OCT-1997; 97US-0062592.

XX PA (GENZ ) GENZYME TRANSGENICS CORP.

XX PI Chen LH, Meade H;

XX DR WPI; 1999-302742/25.

XX DR N-PSDB; AAX25587.

XX PT New modified recombinant nucleic acid sequences useful for producing  
 XX malarial DNA vaccine

XX PS Disclosure; Fig 2; 43pp; English.

XX CC This present sequence comprises a 42 kDa C-terminal portion of  
 CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an  
 CC important target for the development of a vaccine against  
 CC Plasmodium falciparum. The C-terminal end of the sequence is  
 CC modified to include a 6xHis tag. Nucleic acids (see AAX25586 and  
 CC AAX25593) encoding MSP-1-42 have been modified according to a method  
 CC of the invention in order to improve expression in mammalian cells  
 CC and in transgenic animals. The invention provides modified  
 CC recombinant nucleic acid sequences and methods for increasing the  
 CC mRNA levels and protein expression of proteins that are difficult  
 CC to express in cell culture systems, especially mammalian cell  
 CC culture systems or in transgenic animals. The preferred difficult  
 CC protein candidates for expression are those derived from lower  
 CC organisms such as parasites, bacteria and viruses that have DNA  
 CC coding sequences of high AT content or which have mRNA instability  
 CC motifs or rare codons relative to the recombinant expression system  
 CC to be used. The invention allows expression of MSP-1 in the milk  
 CC of transgenic animals, and also provides a DNA vaccine comprising a  
 CC vector containing the altered MSP-1-42 nucleic acid.

XX Sequence 361 AA;

Query Match 10.7%; Score 46; DB 20; Length 361;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-37;  
 Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 NSGCFRHLDERECKLLNYKQEGDKCVENPPTCENNNGCCDADA 397  
 ID AAY05834 standard; Protein; 376 AA.

DB 277 NSGCFRHLDERECKLLNYKQEGDKCVENPPTCENNNGCCDADA 322

## RESULT 30

AA09374  
 ID AAY09374 standard; Protein; 376 AA.

XX AC AAY09374;

XX 31-AUG-1999 (first entry)

DE Modified merozoite surface protein MSP-1-42.

XX MSP-1; merozoite surface protein; malaria; vaccine;  
 KW protein engineering; protein expression; codon usage;  
 KW transgenic animal; mutant.

XX Plasmodium falciparum.

OS Synthetic.

XX Key Location/Qualifiers

FT Peptide 1..15

FT Peptide /note= "beta-casein signal peptide"

FT Peptide 371..376

FT /note= "6xHis tag"

FT Misc-difference 197

FT /note= "Asn in native MSP-1-42 (N181Q mutation)"

FT Misc-difference 278

FT /note= "Asn in native MSP-1-42 (N262Q mutation)"

XX WO9920774-A2.

XX 29-APR-1999.

XX 20-OCT-1998; 98WO-US22226.

XX 15-MAY-1998; 98US-0085649.

XX 20-OCT-1997; 97US-0062592.

XX (GENZ ) GENZYME TRANSGENICS CORP.

XX Chen LH, Meade H;

XX WPI; 1999-288313/24.

XX P-PSDB; AAX56008.

XX Modified malarial protein for use in anti-malarial vaccines

XX Example; Fig 11; 35pp; English.

XX The present sequence represents a 42 kDa C-terminal portion of  
 CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an  
 CC important target for the development of a vaccine against  
 CC Plasmodium falciparum. The sequence has been modified to include  
 CC an N-terminal beta-casein signal peptide and an N-terminal 6xHis  
 CC tag. In addition, N181Q and N262Q mutations have been introduced to  
 CC eliminate 2 N-glycosylation sites. These modifications allow the  
 CC MSP-1-42 protein to be expressed in the milk of transgenic mice.

XX Sequence 376 AA;

Query Match 10.7%; Score 46; DB 20; Length 376;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-37;  
 Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 NSGCFRHLDERECKLLNYKQEGDKCVENPPTCENNNGCCDADA 397  
 ID AAY05834 standard; Protein; 376 AA.

DB 292 NSGCFRHLDERECKLLNYKQEGDKCVENPPTCENNNGCCDADA 337

## RESULT 31

AA05834  
 ID AAY05834 standard; Protein; 376 AA.

XX AC AAY05834;

XX 02-AUG-1999 (first entry)

XX Modified merozoite surface protein MSP-1-42.

XX MSP-1; merozoite surface protein; malaria; vaccine;  
 KW protein engineering; protein expression; codon usage;  
 KW transgenic animal; mutant.

XX Plasmodium falciparum.

OS Synthetic.

XX Key Location/Qualifiers

FT Peptide 1..15

FT Peptide /note= "beta-casein signal peptide"

FT Peptide 371..376

FT /note= "6xHis tag"

FT Misc-difference 197

FT /note= "Asn in native MSP-1-42 (N181Q mutation)"

FT Misc-difference 278

FT /note= "Asn in native MSP-1-42 (N262Q mutation)"

XX WO9920766-A2.

XX 29-APR-1999.

XX 20-OCT-1998; 98WO-US22225.

XX 15-MAY-1998; 98US-0085649.

XX 20-OCT-1997; 97US-0062592.

XX (GENZ ) GENZYME TRANSGENICS CORP.

XX Chen LH, Meade H;

XX WPI; 1999-302742/25.

XX N-PSDB; AAX25593.

XX New modified recombinant nucleic acid sequences useful for producing

XX malarial DNA vaccine

XX Example; Fig 11; 43pp; English.

XX The present sequence represents a 42 kDa C-terminal portion of  
 CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an  
 CC important target for the development of a vaccine against  
 CC Plasmodium falciparum. The sequence has been modified to include  
 CC an N-terminal beta-casein signal peptide and an N-terminal 6xHis  
 CC tag. In addition, N181Q and N262Q mutations have been introduced to  
 CC eliminate 2 N-glycosylation sites. These modifications allow the  
 CC MSP-1-42 protein to be expressed in the milk of transgenic mice.  
 CC The invention generally relates to modified recombinant nucleic  
 CC acid sequences and methods for increasing the mRNA levels and  
 CC protein expression of proteins that are difficult to express in  
 CC cell culture systems, mammalian cell culture systems or in  
 CC transgenic animals. The preferred difficult protein candidates for  
 CC expression are those derived from lower organisms such as parasites,  
 CC bacteria and viruses that have DNA coding sequences of high AT  
 CC content or which have mRNA instability motifs or rare codons  
 CC relative to the recombinant expression system to be used.

XX Sequence 376 AA;

Query Match 10.7%; Score 46; DB 20; Length 376;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-37;

Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 352 NSGCFRHLDERECKCLLNKQEGDKCVENPNTCNENNGGCDADA 397  
 |||||  
 Db 292 NSGCFRHLDERECKCLLNKQEGDKCVENPNTCNENNGGCDADA 337

RESULT 32  
 AAW84315  
 ID AAW84315 standard; Protein; 408 AA.  
 AC AAW84315;  
 XX  
 DT 22-MAR-1999 (first entry)  
 XX  
 DE TrxA-rabbit tissue factor fusion protein.  
 XX  
 KW Thioredoxin A protein; rabbit tissue factor; RTF; trxA;  
 KW blood coagulation cascade; fusion protein; chimeric;  
 KW standard blood clotting test; blood disorder; blood coagulation;  
 KW prothrombin time test.  
 XX  
 OS Chimeric - Oryctolagus cuniculus.  
 OS Chimeric - Escherichia coli.  
 OS Chimeric - Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Protein 1..109  
 FT /note= "Thioredoxin A protein"  
 FT Cleavage-site 126..131  
 FT /note= "thrombin cleavage site"  
 FT Peptide 134..148  
 FT /note= "S-peptide marker"  
 FT Cleavage-site 154..158  
 FT /note= "enterkinase cleavage site"  
 FT Protein 169..408  
 FT /note= "truncated rabbit tissue factor"  
 XX  
 PN US5858724-A.  
 XX  
 PD 12-JAN-1999.  
 XX  
 PF 16-JUL-1996; 96US-0683007.  
 XX  
 PR 16-JUL-1996; 96US-0683007.  
 XX  
 PA (PELF-) PEL FREEZ.  
 XX  
 PI Domanico MJ, Kroecker W, Novy RE, Yaeger KW;  
 XX  
 DR WPI; 1999-119880/10.  
 DR N-PSDB; AAV63788.  
 XX  
 PT Fusion protein containing; truncated rabbit tissue factor;  
 PT thioredoxin A protein and; a tag for affinity purification - useful  
 PT in blood clotting assays and can be produced as a soluble protein in  
 PT the correctly folded form  
 XX  
 PS Claim 11; Columns 13-16; 13pp; English.  
 XX  
 CC The present sequence represents a fusion protein that comprises, in  
 CC an N to C terminal direction, a thioredoxin A (trxA) protein from  
 CC Escherichia coli, a His marker, a S-peptide and a truncated rabbit  
 CC tissue factor (RTF) sequence that includes the extracellular and  
 CC transmembrane domains, but not the cytoplasmic domain, of a native  
 CC mature RTF. RTF can function as part of a blood coagulation cascade.  
 CC The fusion protein is useful in standard blood clotting tests performed  
 CC before surgery to identify disorders of blood coagulation, e.g. the  
 CC prothrombin time test.  
 XX  
 SQ Sequence 408 AA;

Query Match 10.7%; Score 46; DB 20; Length 408;

Best Local Similarity 100.0%; Pred. No. 1.6e-37;  
 Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHHHHHSSGLVPRGSGMKETAATAAKFERQHRMDSPLDGLTDDDDKAMA 46  
 |||||  
 Db 116 MHHHHHSSGLVPRGSGMKETAATAAKFERQHRMDSPLDGLTDDDDKAMA 161

RESULT 33  
 AAW54145  
 ID AAW54145 standard; Protein; 1639 AA.  
 AC AAW54145;  
 XX  
 DT 23-SEP-1998 (first entry)  
 XX  
 DE P. falciparum synthetic gpl90 protein.  
 XX  
 KW gpl90; malaria; MSP-1; merozoite surface protein; stability; vaccine;  
 KW monoclonal antibody; passive immunisation; parasite.  
 XX  
 OS Plasmodium falciparum.  
 OS Synthetic.  
 XX  
 PN WO9814583-A2.  
 XX  
 PD 09-APR-1998.  
 XX  
 PF 02-OCT-1997; 97WO-EP05441.  
 XX  
 PR 02-OCT-1996; 96DE-4040817.  
 XX  
 PA (BUJA/) BUJARD H.  
 XX  
 PI Bujard H, Pan W, Tolle R;  
 XX  
 DR WPI; 1998-240088/21.  
 DR N-PSDB; AAV21451, AAV35363.  
 XX  
 PT Recombinant production of complete gp190/MSP-1 Plasmodium surface  
 PT protein - useful in anti-malaria vaccines, also stabilising genes by  
 PT -reducing their AT content  
 XX  
 PS Example 1; Fig 3c; 48pp; German.  
 XX  
 CC This sequence represents a modified Plasmodium falciparum gp190/MSP-1  
 CC (merozoite surface) protein. The gene encoding this protein has been  
 CC stabilised by reducing the AT content of the nucleotide sequence. Such a  
 CC protein is useful in vaccines against malaria or for producing monoclonal  
 CC antibodies (for passive immunisation). The complete gp190 protein can now  
 CC be produced outside the parasite and has, at least over extended regions,  
 CC the native pattern of folding. Larger amounts of the protein can be  
 CC produced recombinantly than would be possible using the parasites as  
 CC source.  
 XX  
 SQ Sequence 1639 AA;

Query Match 10.7%; Score 46; DB 19; Length 1639;  
 Best Local Similarity 100.0%; Pred. No. 5.9e-37;  
 Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 352 NSGCFRHLDERECKCLLNKQEGDKCVENPNTCNENNGGCDADA 397  
 |||||  
 Db 1540 NSGCFRHLDERECKCLLNKQEGDKCVENPNTCNENNGGCDADA 1585

RESULT 34  
 ABB80661  
 ID ABB80661 standard; peptide; 59 AA.  
 XX  
 AC ABB80661;  
 XX  
 DT 15-JUL-2002 (first entry)



PN WO9317107-A.  
XX 02-SEP-1993.  
XX 22-FEB-1993; 93WO-GB00367.  
XX 22-FEB-1992; 92GB-0003821.  
XX (MEDI-) MEDICAL RES COUNCIL.  
XX Blackman MJ, Chappel JA, Holder AA;  
PI WPT; 1993-288413/36.  
DR  
XX Allelic variants of epidermal growth factor 1- or 2-like domains - of  
PT merozoite surface protein 1, produced recombinantly for malaria  
PT vaccines  
XX  
PS Claim 1; Fig 1a; 35pp; English.  
XX  
CC The sequence is that of an allelic variant of a merozoite surface  
CC protein-1 epidermal growth factor (EGF) 1 like domain. It may be  
CC used alone or as part of a fusion protein of EGF-1-like and  
CC EGF-2-like domains in vaccines against malaria. When expressed  
CC recombinantly it is produced in a form indistinguishable from  
CC that in the native protein.  
XX  
SQ Sequence 49 AA;  
Query Match 7.98; Score 34; DB 14; Length 49;  
Best Local Similarity 100.0%; Pred. No. 2.7e-26;  
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 352 NSGCFRHLDERECKCLLNKQEGDKVCVENPNT 385  
DB 16 NSGCFRHLDERECKCLLNKQEGDKVCVENPNT 49  
RESULT 37  
AAB51128  
ID AAB51128 standard; Protein: 55 AA.  
AC AAB51128;  
XX  
DT 20-MAR-2001 (first entry)  
XX  
DE Human mamaglobin N-terminal fragment SEQ ID NO:28.  
XX  
KW Human; mamaglobin; breast cancer; detection; diagnosis; antibody;  
KW vaccine; cytostatic; antimamaglobin.  
XX  
OS Homo sapiens.  
XX  
PN WO200073338-A1.  
XX  
PD 07-DEC-2000.  
XX  
PF 26-MAY-2000; 2000WO-US14845.  
XX  
PR 28-MAY-1999; 99US-0136528.  
PR 01-JUN-1999; 99US-0137048.  
XX  
PA (CORI-) CORIXA CORP.  
PA (HEND/) HENDRICKSON R C.  
PA (HOUG/) HOUGHTON R L.  
PA (REED/) REED S G.  
XX  
PI Fanger GR;  
XX  
DR WPT; 2001-049928/06.  
XX  
PT Polypeptide comprising at least seven consecutive amino acid residues  
PT of human mamaglobin, useful in the treatment and detection of breast

PT cancer -  
XX  
PS  
XX Example 1; Fig 2; 109pp; English.  
CC  
CC The present invention describes human mamaglobin peptides (I)  
CC comprising at least 7 consecutive residues. Also described are: (1) a  
CC vaccine comprising (I) with an immunostimulant which is an adjuvant;  
CC (2) an isolated antibody (Ab1) or its antigen-binding fragment, which  
CC specifically binds to a mamaglobin epitope having the sequence of  
CC pro2-3; (3) an isolated antibody (Ab2) or its antigen-binding fragment  
CC that specifically binds to glycosylated mamaglobin; (4) a method for  
CC inhibiting the development of breast cancer in a patient, comprising  
CC administering (I) or Ab1 or Ab2; and (5) a method (M1) for determining  
CC the presence or absence of breast cancer in a patient. (I) has  
CC cytostatic activity. The polypeptides and antibodies are used in  
CC vaccines for the prevention and treatment of breast cancer. They are  
CC also used for diagnosis and monitoring of breast cancer. The present  
CC sequence represents a human mamaglobin N-terminal amino acid sequence  
CC fragment, which is used in an example from the present invention.  
XX  
SQ Sequence 55 AA;  
Query Match 7.48; Score 32; DB 22; Length 55;  
Best Local Similarity 100.0%; Pred. No. 3.1e-24;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 15 GSGMKETAARAFERQHMDSPDLGTDDDDKAMA 46  
DB 1 GSGMKETAARAFERQHMDSPDLGTDDDDKAMA 32  
RESULT 38  
AAY39952  
ID AAY39952 standard; Protein: 382 AA.  
XX  
AC AAY39952;  
XX  
DT 13-DEC-1999 (first entry)  
XX  
DE Gaussia luciferase fusion protein sequence.  
XX  
KW Luciferase; green fluorescent protein; GFP; screening assay; diagnosis;  
KW bioluminescence-generating system; toy; cosmetic; fairy dust; beverage;  
KW body paint; squirt gun; balloon; slimy play material; soap; toothpaste;  
KW fusion protein.  
XX  
OS Gaussia sp.  
XX  
PN WO9949019-A2.  
XX  
PD 30-SEP-1999.  
XX  
PF 26-MAR-1999; 99WO-US06698.  
XX  
PR 27-MAR-1998; 98US-0079624.  
PR 15-JUN-1998; 98US-0089367.  
PR 01-OCT-1998; 98US-0102939.  
XX  
PA (PROL-) PROLUME LTD.  
PA (BRYA/) BRYAN B J.  
XX  
PI Bryan BJ, Szent-Gyorgyi C;  
XX  
DR WPT; 1999-580443/49.  
DR N-PSDB; AAZ27550.  
XX  
PT New isolated Renilla mulleri, Gaussia and Pleuromamma luciferase and  
PT Renilla and Ptilosarcus green fluorescent protein nucleic acids -  
XX  
PS Disclosure; Page 222-223; 233pp; English.  
XX  
CC This sequence represents a luciferase of the invention. The invention  
CC relates to Renilla mulleri, Gaussia and Pleuromamma luciferase and



CC Renilla and *Ptilosarcus* green fluorescent protein (GFP) nucleic acids and  
 CC proteins. The luciferases and GFPs can be used in  
 CC bioluminescence-generating systems, assays, screening methods, diagnostic  
 CC method and articles of manufacture. They can be expressed using  
 CC e.g. bacterial, yeast, fungal, plant, insect or animal cells. The Renilla  
 CC mulleri, *Gaussia* and *Pleuromamma* luciferase or Renilla or *Ptilosarcus*  
 CC GFPs can be used in e.g. toys, cosmetics, fountains, personal care items,  
 CC fairy dust, beverages, soft drinks, foods, textile products, bubbles,  
 CC balloons, personal items, dentifrices, soaps, body paints, bubble bath,  
 CC ink or paper products. In particular, they can be used in e.g. squirt  
 CC guns, pellet guns, finger paints, foot bags, greeting cards, slimy play  
 CC material, clothing, bubble making toys, bath powders, cosmetics, body  
 CC lotions, gels, body powders, body creams, bath powders, mouthwashes,  
 CC soaps, body paints, bubble bath, inks, wrapping paper, gelatins, icings,  
 CC frostings, greeting cards, beer, wine, champagne, soft drinks, ice cubes,  
 CC ice, dry ice or fountains. The nucleic acids can also be used to produce  
 CC transgenic fish and plants.

XX Sequence 382 AA;

Query Match 6.3%; Score 27; DB 20; Length 382;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-18;  
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17 GKKETAATAAKFERQHMDSPDLGTDDDDK 43  
 ||||||||||||||||||||||||||||||||  
 Db 171 GKKETAATAAKFERQHMDSPDLGTDDDDK 197

RESULT 39  
 AAEL13383

ID AAEL13383 standard; Protein; 382 AA.

AC AAEL13383;

DT 12-FEB-2002 (first entry)

XX *Gaussia* species CBD-luciferase fusion protein.

XX Green fluorescent protein; GFP; bioluminescence generating system; toy;  
 KW luciferase; finger paint; slimy play material; fishing lure; sparkler;  
 KW doll; balloon; personal care item; cosmetic; bath powder; body cream;  
 KW tooth paste; mouth wash; soap; body paint; ornamental transgenic plant;  
 KW bubble bath; fountain; fairy dust; food; gelatin; icing; frosting; beer;  
 KW wine; champagne; milk; soft drink; ice cube; ice; foot bag; beverage;  
 KW dry ice; textile; clothing; paper product; greeting card; wrapping paper;  
 KW fusion protein.

XX *Gaussia* sp.

PN W020016824-A2.

XX 20-SEP-2001.

XX 15-MAR-2001; 2001WO-US08277.

XX 15-MAR-2000; 2000US-189691P.

XX (PROL-) PROLUME LTD.

PA (BRYA/) BRYAN B J.

XX Bryan BJ, Szent-Gyorgyi C, Szczepaniak W;

XX WPI; 2002-010561/01.

DR N-PSDB; AAD22201.

XX Nucleic acids encoding Renilla reniformis green fluorescent proteins,  
 PT useful in diagnostic bioluminescence procedures -

XX Disclosure; Page 163-164; 175pp; English.

XX The patent discloses sea pansy (*Renilla reniformis*) green fluorescent  
 CC proteins (GFP) and their corresponding polynucleotides. The invention

CC also relates to sequences of the bioluminescence generating system  
 CC (e.g. luciferase). *R. reniformis* GFP are used in diagnostic methods  
 CC and in the production of novelty items such as toys (e.g. squirt gun,  
 CC pellet gun, toy "Halloween" eggs, toy cigarettes, board/card game  
 CC toys), finger paints, slimy play material, bubbles in bubble making  
 CC toys, fishing lures, dolls, sparklers, magic wand toys, balloons,  
 CC personal care item (e.g. cosmetic, bath powders, body creams, tooth  
 CC pastes, mouth wash, soaps, body paints, bubble bath), ornamental  
 CC transgenic plants, fountain, fairy dust, food (gelatins, icings,  
 CC frosting, beer, wine, champagne, milk, soft drinks, ice cubes, ice,  
 CC dry ice, beverage), textile (foot bag, clothing) and/or paper product  
 CC (greeting cards, wrapping paper). The present sequence is CBD-luciferase  
 CC fusion protein from *Gaussia* species.

XX Sequence 382 AA;

Query Match 6.3%; Score 27; DB 23; Length 382;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-18;  
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17 GKKETAATAAKFERQHMDSPDLGTDDDDK 43  
 ||||||||||||||||||||||||||||||||  
 Db 171 GKKETAATAAKFERQHMDSPDLGTDDDDK 197

RESULT 40

AAAR41356

ID AAR41356 standard; peptide; 54 AA.

AC AAR41356;

DT 04-MAR-1994 (first entry)

XX MSP1EGF2A EGF2-like domain variant.

XX Epidermal growth factor 1; merozoite surface protein 1; malaria;  
 KW vaccine.

XX Plasmodium yoelii.

PH Key Location/Qualifiers

FT Cleavage-site 1

FT /note= "introduced to facilitate cleavage  
 from recombinant protein"

FT WO9317107-A.

XX 02-SEP-1993.

XX 22-FEB-1993; 93WO-GB00367.

XX 22-FEB-1992; 92GB-0003821.

XX (MEDI-) MEDICAL RES COUNCIL.

XX Blackman MJ, Chappel JA, Holder AA;

XX WPI; 1993-288413/36.

XX Allelic variants of epidermal growth factor 1- or 2-like domains - of  
 PT merozoite surface protein 1, produced recombinantly for malaria  
 PT vaccines

XX Claim 2; Fig 2a; 35pp; English.

XX The sequence is that of an allelic variant of a merozoite surface  
 CC protein-1 epidermal growth factor (EGF) 2-like domain. It may be  
 CC used alone or as part of a fusion protein of EGF-1-like and  
 CC EGF-2-like domains in vaccines against malaria. When expressed  
 CC recombinantly it is produced in a form indistinguishable from  
 CC that in the native protein.

XX Sequence 54 AA;

QY 409 KKITCECTKPDSPYPLFDGIFCSS 431  
 |||||  
 Db 30 KKITCECTKPDSPYPLFDGIFCSS 52

Search completed: May 12, 2003, 10:27:22  
Job time : 62 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 10:26:03 ; Search time 18 Seconds  
(without alignments)  
704.516 Million cell updates/sec

Title: US-10-057-531A-2

Perfect score: 431

Sequence: 1 MHHHHHSSGLVPRGSMKE.....TCECTKPDSPFLDGFICSS 431

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 2942292 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued\_Patents\_AA.\*

- 1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	231	53.6	394	4	US-08-195-705-2
2	135	31.3	394	4	US-08-195-705-4
3	50	11.6	53	1	US-08-290-919-4
4	50	11.6	167	4	US-09-741-243C-4
5	50	11.6	467	4	US-09-513-442-2
6	50	11.6	551	4	US-09-741-243C-2
7	50	11.6	824	4	US-09-626-589-3
8	47	10.9	48	1	US-08-290-919-2
9	46	10.7	408	2	US-08-683-007A-2
10	45	10.4	377	4	US-08-195-705-5
11	34	7.9	48	1	US-08-290-919-1
12	34	7.9	48	1	US-08-290-919-12
13	27	6.3	209	4	US-09-166-966E-11
14	27	6.3	382	4	US-09-277-716-22
15	27	6.3	382	4	US-09-609-161B-22
16	24	5.6	159	4	US-09-166-966E-8
17	24	5.6	375	4	US-08-195-705-3
18	23	5.3	53	1	US-08-290-919-3
19	15	3.5	15	1	US-08-470-837-24
20	15	3.5	15	2	US-08-639-806-1
21	15	3.5	15	4	US-08-868-452-24
22	15	3.5	20	1	US-08-373-134D-3
23	15	3.5	20	2	US-09-114-637-3
24	15	3.5	20	3	US-08-485-942A-99
25	15	3.5	20	3	US-08-488-214A-99
26	15	3.5	20	4	US-08-438-431A-99
27	15	3.5	20	4	US-09-638-202A-114

28	15	3.5	21	2	US-08-432-871C-46	Sequence 46, Appl
29	15	3.5	21	2	US-08-712-878-6	Sequence 6, Appl
30	15	3.5	21	3	US-08-485-942A-98	Sequence 98, Appl
31	15	3.5	21	3	US-08-488-214A-98	Sequence 98, Appl
32	15	3.5	21	3	US-08-488-208A-98	Sequence 98, Appl
33	15	3.5	21	4	US-08-483-211A-98	Sequence 98, Appl
34	15	3.5	21	4	US-08-488-223A-98	Sequence 98, Appl
35	15	3.5	21	4	US-08-438-431A-98	Sequence 98, Appl
36	15	3.5	21	4	US-09-270-956-46	Sequence 46, Appl
37	15	3.5	23	4	US-08-761-483-7	Sequence 7, Appl
38	15	3.5	43	2	US-08-347-563A-12	Sequence 12, Appl
39	15	3.5	43	3	US-08-485-942A-12	Sequence 12, Appl
40	15	3.5	43	3	US-08-488-214A-12	Sequence 12, Appl
41	15	3.5	43	3	US-08-488-208A-12	Sequence 12, Appl
42	15	3.5	43	4	US-08-483-211A-12	Sequence 12, Appl
43	15	3.5	43	4	US-08-488-223A-12	Sequence 12, Appl
44	15	3.5	43	4	US-08-438-431A-12	Sequence 12, Appl
45	15	3.5	110	1	US-08-434-705B-15	Sequence 15, Appl
46	15	3.5	110	2	US-09-086-201-15	Sequence 15, Appl
47	15	3.5	124	1	US-08-184-604-2	Sequence 2, Appl
48	15	3.5	124	2	US-08-891-848-18	Sequence 18, Appl
49	15	3.5	124	3	US-08-875-811-13	Sequence 13, Appl
50	15	3.5	124	6	5171845-11	Patent No. 5171845
51	15	3.5	144	4	US-09-230-421-3	Sequence 3, Appl
52	15	3.5	153	4	US-09-675-922-2	Sequence 2, Appl
53	15	3.5	157	4	US-09-675-922-4	Sequence 4, Appl
54	15	3.5	163	4	US-09-675-922-6	Sequence 6, Appl
55	15	3.5	167	4	US-09-675-922-8	Sequence 8, Appl
56	15	3.5	215	4	US-09-198-723A-46	Sequence 46, Appl
57	15	3.5	215	4	US-09-198-723A-50	Sequence 50, Appl
58	15	3.5	215	4	US-09-198-723A-53	Sequence 53, Appl
59	15	3.5	215	4	US-09-198-723A-57	Sequence 57, Appl
60	15	3.5	216	4	US-09-198-723A-1	Sequence 1, Appl
61	15	3.5	216	4	US-09-198-723A-2	Sequence 2, Appl
62	15	3.5	216	4	US-09-198-723A-3	Sequence 3, Appl
63	15	3.5	216	4	US-09-198-723A-4	Sequence 4, Appl
64	15	3.5	216	4	US-09-198-723A-5	Sequence 5, Appl
65	15	3.5	216	4	US-09-198-723A-6	Sequence 6, Appl
66	15	3.5	216	4	US-09-198-723A-7	Sequence 7, Appl
67	15	3.5	216	4	US-09-198-723A-8	Sequence 8, Appl
68	15	3.5	216	4	US-09-198-723A-9	Sequence 9, Appl
69	15	3.5	216	4	US-09-198-723A-10	Sequence 10, Appl
70	15	3.5	288	4	US-09-273-839A-8	Sequence 8, Appl
71	15	3.5	379	4	US-09-457-040B-36	Sequence 36, Appl
72	15	3.5	431	4	US-09-376-689-4	Sequence 4, Appl
73	15	3.5	666	4	US-09-198-723A-11	Sequence 11, Appl
74	15	3.5	666	4	US-09-198-723A-12	Sequence 12, Appl
75	15	3.5	666	4	US-09-198-723A-13	Sequence 13, Appl
76	15	3.5	666	4	US-09-198-723A-14	Sequence 14, Appl
77	15	3.5	666	4	US-09-198-723A-15	Sequence 15, Appl
78	15	3.5	666	4	US-09-198-723A-16	Sequence 16, Appl
79	15	3.5	666	4	US-09-198-723A-17	Sequence 17, Appl
80	15	3.5	666	4	US-09-198-723A-18	Sequence 18, Appl
81	15	3.5	672	4	US-09-198-723A-19	Sequence 19, Appl
82	15	3.5	712	4	US-09-198-723A-20	Sequence 20, Appl
83	15	3.5	711	3	US-08-946-475-9	Sequence 9, Appl
84	15	3.5	711	4	US-09-340-479-9	Sequence 9, Appl
85	13	3.0	15	3	US-08-350-551-2	Sequence 2, Appl
86	13	3.0	15	4	US-09-561-736-2	Sequence 2, Appl
87	13	3.0	17	3	US-08-946-475-10	Sequence 10, Appl
88	13	3.0	17	4	US-09-340-479-10	Sequence 10, Appl
89	12	2.8	18	1	US-07-988-754-6	Sequence 6, Appl
90	12	2.8	124	2	US-08-891-848-17	Sequence 17, Appl
91	12	2.8	124	3	US-08-875-811-12	Sequence 12, Appl
92	10	2.3	19	1	US-08-737-757-15	Sequence 15, Appl
93	9	2.1	16	4	US-09-506-768-15	Sequence 15, Appl
94	9	2.1	20	1	US-09-186-489-6	Sequence 6, Appl
95	9	2.1	21	2	US-08-651-818A-21	Sequence 21, Appl
96	9	2.1	21	3	US-09-184-826-21	Sequence 21, Appl
97	9	2.1	23	1	US-08-480-604A-24	Sequence 24, Appl
98	9	2.1	23	2	US-08-405-496A-24	Sequence 24, Appl
99	9	2.1	23	3	US-08-915-136-24	Sequence 24, Appl
100	9	2.1	24	3	US-08-584-031-8	Sequence 8, Appl







539	6	1.4	26	2	US-08-657-641-12	Sequence 12, Appl	612	77	4	US-09-643-597-361	Sequence 361, App
540	6	1.4	26	3	US-08-470-397-2	Sequence 2, Appl	613	79	2	US-08-448-418-97	Sequence 97, Appl
541	6	1.4	26	3	US-08-895-707-21	Sequence 21, Appl	614	80	4	US-09-174-943-4	Sequence 4, Appl
542	6	1.4	26	5	PCT-US94-07233-22	Sequence 22, Appl	615	81	2	US-08-717-169-15	Sequence 15, Appl
543	6	1.4	27	4	US-09-411-329C-22	Sequence 22, Appl	616	81	2	US-08-717-169-15	Sequence 15, Appl
544	6	1.4	27	4	US-09-411-329C-26	Sequence 26, Appl	617	85	3	US-09-025-151-26	Sequence 26, Appl
545	6	1.4	28	1	US-08-664-449-37	Sequence 37, Appl	618	85	3	US-08-772-440-32	Sequence 32, Appl
546	6	1.4	28	3	US-08-978-741-14	Sequence 14, Appl	619	85	4	US-09-256-976-82	Sequence 82, Appl
547	6	1.4	28	4	US-09-333-729A-14	Sequence 14, Appl	620	87	1	US-08-160-670A-7	Sequence 7, Appl
548	6	1.4	28	4	US-08-393-985-33	Sequence 33, Appl	621	88	1	US-08-308-086-11	Sequence 11, Appl
549	6	1.4	29	1	PCT-US94-05150-37	Sequence 37, Appl	622	89	1	US-08-167-035-22	Sequence 22, Appl
550	6	1.4	29	5	US-09-039-780A-98	Sequence 98, Appl	623	89	1	US-08-340-812-13	Sequence 13, Appl
551	6	1.4	30	4	US-09-039-780A-100	Sequence 100, Appl	624	89	1	US-08-208-887A-22	Sequence 22, Appl
552	6	1.4	30	4	US-09-039-780A-104	Sequence 104, Appl	625	89	1	US-08-459-064B-13	Sequence 13, Appl
553	6	1.4	31	2	US-08-662-227-37	Sequence 37, Appl	626	89	2	US-08-479-078-16	Sequence 16, Appl
554	6	1.4	31	4	US-09-017-947-37	Sequence 37, Appl	627	89	2	US-08-460-421A-23	Sequence 13, Appl
555	6	1.4	32	1	US-08-323-445A-17	Sequence 17, Appl	628	89	2	US-08-539-005-22	Sequence 22, Appl
556	6	1.4	32	1	US-08-515-903A-17	Sequence 17, Appl	629	89	4	US-08-717-169-7	Sequence 7, Appl
557	6	1.4	32	3	US-08-737-336-1	Sequence 1, Appl	630	89	4	US-09-280-598-24	Sequence 24, Appl
558	6	1.4	32	4	US-08-411-760-7	Sequence 7, Appl	631	94	4	US-09-256-976-95	Sequence 95, Appl
559	6	1.4	32	4	US-08-411-760-9	Sequence 9, Appl	632	96	1	US-07-807-529A-39	Sequence 39, Appl
560	6	1.4	32	4	US-08-411-760-10	Sequence 10, Appl	633	96	3	US-08-430-944D-103	Sequence 103, App
561	6	1.4	32	5	PCT-US95-12840-17	Sequence 17, Appl	634	96	4	US-08-431-184-103	Sequence 103, App
562	6	1.4	34	2	US-08-468-558-8	Sequence 8, Appl	635	97	3	US-09-605-785-827	Sequence 827, App
563	6	1.4	34	2	US-08-468-558-9	Sequence 9, Appl	636	97	3	US-08-816-977-23	Sequence 23, Appl
564	6	1.4	34	4	US-08-676-444-8	Sequence 8, Appl	637	97	3	US-08-816-977-27	Sequence 27, Appl
565	6	1.4	34	4	US-08-676-444-9	Sequence 9, Appl	638	106	1	US-09-174-943-2	Sequence 4, Appl
566	6	1.4	34	4	US-09-314-268-118	Sequence 118, App	639	106	1	US-07-623-611-4	Sequence 7, Appl
567	6	1.4	37	4	US-09-039-780A-97	Sequence 97, Appl	640	106	1	US-07-623-611-7	Sequence 7, Appl
568	6	1.4	39	2	US-08-679-865-39	Sequence 39, Appl	641	106	4	US-08-290-919-11	Sequence 11, Appl
569	6	1.4	39	2	US-08-680-876-39	Sequence 39, Appl	642	106	4	US-09-113-977C-47	Sequence 47, Appl
570	6	1.4	39	4	US-09-263-975-39	Sequence 39, Appl	643	106	5	PCT-US91-09108-4	Sequence 4, Appl
571	6	1.4	39	4	US-09-450-072-78	Sequence 78, Appl	644	106	5	PCT-US91-09108-7	Sequence 7, Appl
572	6	1.4	39	4	US-09-351-348-78	Sequence 78, Appl	645	107	3	US-08-722-349-6	Sequence 6, Appl
573	6	1.4	40	2	US-08-273-146-39	Sequence 39, Appl	646	107	3	US-09-204-328-6	Sequence 6, Appl
574	6	1.4	40	2	US-08-664-449-39	Sequence 39, Appl	647	108	2	US-08-468-819-74	Sequence 74, Appl
575	6	1.4	42	4	US-09-309-382-12	Sequence 12, Appl	648	108	2	US-08-468-819-76	Sequence 76, Appl
576	6	1.4	42	4	US-08-273-146-41	Sequence 41, Appl	649	120	1	US-08-347-492B-2	Sequence 2, Appl
577	6	1.4	43	2	US-08-652-816A-43	Sequence 43, Appl	650	120	2	US-08-798-143-2	Sequence 2, Appl
578	6	1.4	44	2	US-08-652-507-6	Sequence 6, Appl	651	120	4	US-08-341-560B-2	Sequence 2, Appl
579	6	1.4	44	4	US-08-960-190A-34	Sequence 34, Appl	652	120	4	US-08-353-940-2	Sequence 2, Appl
580	6	1.4	44	4	US-09-039-780A-99	Sequence 99, Appl	653	120	5	PCT-US93-03895-2	Sequence 2, Appl
581	6	1.4	45	4	US-08-960-190A-33	Sequence 33, Appl	654	120	5	PCT-US95-15484-2	Sequence 2, Appl
582	6	1.4	48	1	US-08-457-245-21	Sequence 21, Appl	655	121	4	US-08-341-560B-4	Sequence 4, Appl
583	6	1.4	50	4	US-09-360-237-53	Sequence 53, Appl	656	121	4	US-08-353-940-4	Sequence 4, Appl
584	6	1.4	52	3	US-08-978-741-15	Sequence 15, Appl	657	121	5	PCT-US93-03895-4	Sequence 4, Appl
585	6	1.4	52	4	US-09-333-729A-15	Sequence 15, Appl	658	124	1	US-08-434-705B-16	Sequence 16, Appl
586	6	1.4	55	1	US-08-466-033-273	Sequence 273, App	659	124	2	US-09-086-201-16	Sequence 16, Appl
587	6	1.4	55	2	US-08-444-733-273	Sequence 273, App	660	124	2	US-09-199-242E-3	Sequence 3, Appl
588	6	1.4	55	2	US-08-464-134-273	Sequence 273, App	661	124	4	US-09-199-242E-5	Sequence 5, Appl
589	6	1.4	55	2	US-08-461-361-273	Sequence 273, App	662	124	4	US-09-042-071-39	Sequence 39, Appl
590	6	1.4	55	2	US-08-485-910-273	Sequence 273, App	663	124	4	US-08-268-348A-12	Sequence 12, Appl
591	6	1.4	57	2	US-08-652-816A-41	Sequence 41, Appl	664	126	6	5171845-12	Patent No. 5171845
592	6	1.4	57	2	US-08-958-201-18	Sequence 18, Appl	665	127	3	US-08-467-023-188	Sequence 188, App
593	6	1.4	63	4	US-09-076-804-4	Sequence 4, Appl	666	127	3	US-08-467-023-189	Sequence 189, App
594	6	1.4	63	4	US-09-134-001C-2901	Sequence 2901, Ap	667	127	3	US-08-467-023-190	Sequence 190, App
595	6	1.4	66	3	US-08-782-480-26	Sequence 26, Appl	668	128	3	US-08-467-023-187	Sequence 187, App
596	6	1.4	66	4	US-08-954-211-26	Sequence 26, Appl	669	130	4	US-08-240-839-4	Sequence 4, Appl
597	6	1.4	66	4	US-09-005-167A-26	Sequence 26, Appl	670	131	4	US-09-367-309A-2	Sequence 2, Appl
598	6	1.4	66	4	US-09-176-741B-26	Sequence 26, Appl	671	133	1	US-08-268-348A-8	Sequence 8, Appl
599	6	1.4	67	2	US-08-142-551B-9	Sequence 9, Appl	672	133	1	US-08-268-348A-10	Sequence 10, Appl
600	6	1.4	67	3	US-09-268-070-3	Sequence 3, Appl	673	136	5	PCT-US95-07171-2	Sequence 2, Appl
601	6	1.4	68	3	US-09-268-070-1	Sequence 1, Appl	674	139	4	US-08-856-253-8	Sequence 8, Appl
602	6	1.4	68	4	US-09-174-943-6	Sequence 6, Appl	675	147	4	US-08-122-458D-9	Sequence 9, Appl
603	6	1.4	69	2	US-08-687-865A-21	Sequence 21, Appl	676	149	4	US-09-134-001C-5151	Sequence 5151, Ap
604	6	1.4	69	3	US-09-268-070-2	Sequence 2, Appl	677	151	1	US-08-321-358-2	Sequence 2, Appl
605	6	1.4	71	4	US-09-043-711-21	Sequence 21, Appl	678	157	4	US-08-875-533-46	Sequence 46, Appl
606	6	1.4	71	1	US-08-309-152-4	Sequence 4, Appl	679	157	4	US-09-347-801-6	Sequence 6, Appl
607	6	1.4	71	4	US-09-146-054-9	Sequence 9, Appl	680	158	4	US-08-875-533-47	Sequence 47, Appl
608	6	1.4	71	4	US-09-664-977A-9	Sequence 9, Appl	681	159	3	US-08-991-890-4	Sequence 4, Appl
609	6	1.4	72	4	US-09-042-071-38	Sequence 38, Appl	682	159	4	US-08-856-253-2	Sequence 2, Appl
610	6	1.4	76	3	US-08-956-307B-17	Sequence 17, Appl	683	160	4	US-09-355-700-59	Sequence 59, Appl
611	6	1.4	76	3	US-08-956-307B-18	Sequence 18, Appl	684	171	1	US-08-680-726A-76	Sequence 76, Appl





831	6	1.4	290	2	US-08-903-801-1	Sequence 1, Appli	904	1.4	349	4	US-09-011-769A-64	Sequence 64, Appl
832	6	1.4	291	1	US-09-295-055-1	Sequence 1, Appli	905	1.4	351	2	US-08-701-191A-3	Sequence 3, Appli
833	6	1.4	291	4	US-08-102-757-9	Sequence 9, Appli	906	1.4	352	1	US-08-395-507-1	Sequence 1, Appli
834	6	1.4	292	4	US-09-485-895-10	Sequence 10, Appl	907	1.4	354	2	US-08-216-894-4	Sequence 4, Appli
835	6	1.4	295	1	US-08-149-091-1	Sequence 1, Appli	908	1.4	354	4	US-09-115-746-4	Sequence 4, Appli
836	6	1.4	295	2	US-08-637-761-1	Sequence 1, Appli	909	1.4	357	4	US-09-605-785-826	Sequence 826, App
837	6	1.4	295	5	PCT-US94-12672-1	Sequence 1, Appli	910	1.4	362	1	US-08-395-507-2	Sequence 2, Appli
838	6	1.4	298	1	US-08-149-091-6	Sequence 6, Appli	911	1.4	364	4	US-09-457-040B-34	Sequence 34, Appl
839	6	1.4	298	2	US-08-637-761-6	Sequence 6, Appli	912	1.4	365	4	US-09-457-040B-35	Sequence 35, Appl
840	6	1.4	298	4	US-09-318-661-2	Sequence 2, Appli	913	1.4	366	4	US-09-210-843-2	Sequence 2, Appli
841	6	1.4	298	5	PCT-US94-12672-6	Sequence 6, Appli	914	1.4	366	4	US-09-718-692-4	Sequence 4, Appli
842	6	1.4	300	4	US-09-277-078-2	Sequence 2, Appli	915	1.4	366	4	US-09-718-852-4	Sequence 4, Appli
843	6	1.4	300	4	US-09-561-756-36	Sequence 36, Appl	916	1.4	366	4	US-09-718-815-4	Sequence 4, Appli
844	6	1.4	300	4	US-09-227-721-36	Sequence 36, Appl	917	1.4	368	1	US-08-211-942-17	Sequence 17, Appl
845	6	1.4	301	2	US-08-661-052-14	Sequence 14, Appl	918	1.4	368	4	US-09-433-428D-58	Sequence 58, Appl
846	6	1.4	301	4	US-09-188-082-14	Sequence 14, Appl	919	1.4	368	4	US-09-000-094-20	Sequence 20, Appl
847	6	1.4	301	4	US-09-364-088-14	Sequence 14, Appl	920	1.4	371	4	US-09-483-885-6	Sequence 6, Appli
848	6	1.4	301	4	US-09-102-716-14	Sequence 14, Appl	921	1.4	375	4	US-09-572-191-4	Sequence 4, Appli
849	6	1.4	302	2	US-08-203-532F-4	Sequence 4, Appli	922	1.4	375	4	US-09-000-094-22	Sequence 22, Appl
850	6	1.4	302	2	US-08-893-853-3	Sequence 3, Appli	923	1.4	375	4	US-09-723-262-4	Sequence 4, Appli
851	6	1.4	302	3	US-08-950-860-16	Sequence 16, Appl	924	1.4	375	4	US-09-723-219-4	Sequence 4, Appli
852	6	1.4	302	4	US-09-113-921-3	Sequence 3, Appli	925	1.4	376	3	US-08-751-512-8	Sequence 8, Appli
853	6	1.4	302	4	US-09-078-465-4	Sequence 4, Appli	926	1.4	378	2	US-08-707-237A-104	Sequence 104, App
854	6	1.4	302	5	PCT-US95-01882A-4	Sequence 4, Appli	927	1.4	378	3	US-08-642-246-26	Sequence 26, Appl
855	6	1.4	303	2	US-08-203-532F-2	Sequence 2, Appli	928	1.4	378	4	US-09-451-206-26	Sequence 26, Appl
856	6	1.4	303	3	US-08-928-361B-7	Sequence 7, Appli	929	1.4	378	5	PCT-US96-06229-26	Sequence 26, Appl
857	6	1.4	303	4	US-09-078-465-2	Sequence 2, Appli	930	1.4	382	4	US-09-724-519-4	Sequence 4, Appli
858	6	1.4	303	5	PCT-US95-01882A-2	Sequence 2, Appli	931	1.4	382	4	US-09-592-037-4	Sequence 4, Appli
859	6	1.4	306	4	US-09-386-642-53	Sequence 53, Appl	932	1.4	383	4	US-09-485-885-23	Sequence 23, Appl
860	6	1.4	308	1	US-08-493-568-2	Sequence 2, Appli	933	1.4	384	4	US-09-374-038-4	Sequence 4, Appli
861	6	1.4	308	2	US-08-793-958-2	Sequence 2, Appli	934	1.4	384	4	US-09-658-179-4	Sequence 4, Appli
862	6	1.4	308	4	US-09-804-778A-2	Sequence 2, Appli	935	1.4	386	3	US-08-895-707-7	Sequence 7, Appli
863	6	1.4	308	4	US-09-804-778A-4	Sequence 4, Appli	936	1.4	388	2	US-08-759-581B-9	Sequence 9, Appli
864	6	1.4	309	4	US-09-079-029-9	Sequence 9, Appli	937	1.4	388	4	US-09-304-711-9	Sequence 9, Appli
865	6	1.4	310	4	US-09-079-029-11	Sequence 11, Appl	938	1.4	388	4	US-09-173-281-9	Sequence 9, Appli
866	6	1.4	312	4	US-09-079-029-10	Sequence 10, Appl	939	1.4	390	4	US-09-485-885-14	Sequence 14, Appl
867	6	1.4	315	2	US-08-484-956-91	Sequence 91, Appl	940	1.4	391	2	US-08-244-951A-10	Sequence 10, Appl
868	6	1.4	315	2	US-08-757-653-91	Sequence 91, Appl	941	1.4	391	2	US-08-389-011-23	Sequence 23, Appl
869	6	1.4	315	4	US-08-520-946-91	Sequence 91, Appl	942	1.4	391	3	US-08-403-917A-23	Sequence 23, Appl
870	6	1.4	315	4	US-09-386-653A-9	Sequence 9, Appli	943	1.4	391	3	US-08-348-952A-23	Sequence 23, Appl
871	6	1.4	319	3	US-08-526-136-14	Sequence 14, Appl	944	1.4	391	5	PCT-US95-15696-2	Sequence 2, Appli
872	6	1.4	319	4	US-09-386-642-12	Sequence 12, Appl	945	1.4	397	3	US-08-978-741-6	Sequence 6, Appli
873	6	1.4	320	2	US-08-757-653-163	Sequence 163, App	946	1.4	397	4	US-09-333-729A-7	Sequence 7, Appli
874	6	1.4	320	2	US-08-823-516-61	Sequence 61, Appl	947	1.4	397	4	US-08-327-874A-11	Sequence 11, Appl
875	6	1.4	320	3	US-08-759-038-102	Sequence 102, App	948	1.4	397	5	PCT-US94-09700-11	Sequence 11, Appl
876	6	1.4	320	3	US-08-758-314-102	Sequence 102, App	949	1.4	399	3	US-08-860-519-8	Sequence 8, Appli
877	6	1.4	322	1	US-08-097-828-1	Sequence 1, Appli	950	1.4	400	4	US-09-086-010-2	Sequence 2, Appli
878	6	1.4	322	1	US-08-097-828-3	Sequence 3, Appli	951	1.4	400	4	US-09-572-191-6	Sequence 6, Appli
879	6	1.4	322	1	US-08-480-756-1	Sequence 1, Appli	952	1.4	409	4	US-09-723-219-6	Sequence 6, Appli
880	6	1.4	322	1	US-08-480-756-3	Sequence 3, Appli	953	1.4	409	4	US-08-560-005-9	Sequence 9, Appli
881	6	1.4	322	2	US-08-462-403-1	Sequence 1, Appli	954	1.4	412	3	US-09-418-540-9	Sequence 9, Appli
882	6	1.4	322	2	US-08-462-403-3	Sequence 3, Appli	955	1.4	412	4	US-09-471-396-1	Sequence 1, Appli
883	6	1.4	322	5	PCT-US92-00331-1	Sequence 1, Appli	956	1.4	413	3	US-08-669-408B-10	Sequence 10, Appl
884	6	1.4	322	5	PCT-US92-00331-3	Sequence 3, Appli	957	1.4	413	4	US-09-374-038-2	Sequence 2, Appli
885	6	1.4	322	5	PCT-US93-10419-1	Sequence 3, Appli	958	1.4	413	4	US-09-658-179-2	Sequence 2, Appli
886	6	1.4	322	5	PCT-US93-10419-3	Sequence 3, Appli	959	1.4	414	5	PCT-US92-06840-2	Sequence 2, Appli
887	6	1.4	323	3	US-08-816-977-21	Sequence 21, Appl	960	1.4	415	4	US-09-100-193-2	Sequence 2, Appli
888	6	1.4	326	3	US-08-816-977-25	Sequence 25, Appl	961	1.4	420	2	US-08-845-998-8	Sequence 8, Appli
889	6	1.4	326	4	US-09-230-196-20	Sequence 20, Appl	962	1.4	420	4	US-09-206-537-8	Sequence 8, Appli
890	6	1.4	326	4	US-09-632-947B-1	Sequence 1, Appli	963	1.4	420	4	US-09-430-854-8	Sequence 8, Appli
891	6	1.4	327	4	US-09-386-629-8	Sequence 8, Appli	964	1.4	422	4	US-09-217-228-7	Sequence 7, Appli
892	6	1.4	328	4	US-09-386-642-11	Sequence 11, Appl	965	1.4	423	2	US-08-910-798-2	Sequence 2, Appli
893	6	1.4	331	4	US-09-217-228-6	Sequence 6, Appli	966	1.4	425	1	US-08-313-553-13	Sequence 13, Appl
894	6	1.4	334	4	US-09-523-849-31	Sequence 31, Appl	967	1.4	425	3	US-08-767-993-13	Sequence 13, Appl
895	6	1.4	335	4	US-09-095-117-2	Sequence 2, Appli	968	1.4	426	3	US-08-932-823A-2	Sequence 2, Appli
896	6	1.4	335	4	US-09-095-117-4	Sequence 4, Appli	969	1.4	427	3	US-08-737-248-4	Sequence 4, Appli
897	6	1.4	336	4	US-09-134-001C-4379	Sequence 4379, Ap	970	1.4	427	3	US-08-846-021A-8	Sequence 8, Appli
898	6	1.4	338	2	US-08-447-430A-43	Sequence 43, Appl	971	1.4	430	4	US-09-149-476-707	Sequence 707, App
899	6	1.4	338	2	US-08-602-359A-40	Sequence 40, Appl	972	1.4	432	4	US-09-326-203A-25	Sequence 25, Appl
900	6	1.4	345	4	US-08-856-253-7	Sequence 7, Appli	973	1.4	433	1	US-08-417-492-2	Sequence 2, Appli
901	6	1.4	347	4	US-09-134-001C-3752	Sequence 3752, Ap	974	1.4	434	4	US-09-548-372D-53	Sequence 53, Appl
902	6	1.4	349	4	US-09-011-769A-47	Sequence 47, Appl	975	1.4	434	4	US-09-548-367D-53	Sequence 53, Appl
903	6	1.4	349	4	US-09-011-769A-60	Sequence 60, Appl	976	1.4	434	4		

977 1.4 435 4 US-09-134-001C-4622 Sequence 4622, Ap  
978 1.4 439 4 US-09-433-428D-57 Sequence 57, Appl  
979 1.4 439 4 US-08-506-296B-67 Sequence 67, Appl  
980 1.4 440 3 US-09-259-214-2 Sequence 2, Appli  
981 1.4 440 4 US-09-318-528-2 Sequence 2, Appli  
982 1.4 440 4 US-09-291-931-2 Sequence 2, Appli  
983 1.4 442 4 US-08-506-296B-70 Sequence 70, Appl  
984 1.4 443 4 US-08-506-296B-76 Sequence 76, Appl  
985 1.4 447 4 US-09-370-253-10 Sequence 10, Appl  
986 1.4 447 4 US-08-506-296B-73 Sequence 73, Appl  
987 1.4 448 4 US-09-461-474-8 Sequence 8, Appli  
988 1.4 449 2 US-08-927-394-2 Sequence 2, Appli  
989 1.4 459 4 US-09-548-372B-32 Sequence 32, Appl  
990 1.4 459 4 US-09-548-367D-32 Sequence 32, Appl  
991 1.4 464 2 US-08-553-619B-9 Sequence 9, Appli  
992 1.4 465 4 US-09-000-094-24 Sequence 24, Appl  
993 1.4 474 2 US-08-907-706-6 Sequence 6, Appli  
994 1.4 474 2 US-09-096-982-8 Sequence 8, Appli  
995 1.4 474 2 US-08-553-650A-8 Sequence 8, Appli  
996 1.4 474 4 US-09-556-877-90 Sequence 90, Appl  
997 1.4 474 4 US-09-620-412C-90 Sequence 90, Appl  
998 1.4 477 3 US-08-772-270A-13 Sequence 13, Appl  
999 1.4 479 1 US-08-397-633A-78 Sequence 78, Appl  
1000 1.4 480 5 PCT-US93-04102-2 Sequence 2, Appli

ALIGNMENTS

RESULT 1  
US-08-195-705-2  
; Sequence 2, Application US/08195705  
; Patent No. 6420523  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Sandra  
; APPLICANT: Hul, George  
; APPLICANT: Barr, Philip  
; APPLICANT: Gibson, Helen  
; TITLE OF INVENTION: BACULOVIRUS PRODUCED PLASMIDIUM  
; TITLE OF INVENTION: FALCIPARUM VACCINE  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Davis Hoxie Faithfull Hapgood  
; STREET: 45 Rockefeller Pl.  
; CITY: New York  
; STATE: N.Y.  
; COUNTRY: USA  
; ZIP: 10111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/195,705  
; FILING DATE: 14-FEB-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jacobs, Seth H  
; REGISTRATION NUMBER: 32140  
; REFERENCE/DOCKET NUMBER: 11880A3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-586-1461  
; TELEFAX: 212-586-1461  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 394 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO

; FRAGMENT TYPE: C-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: Plasmodium falciparum  
; STRAIN: falciparum uganda palo alto (FUP)  
; US-08-195-705-2  
Query Match 53.6%; Score 231; DB 4; Length 394;  
Best Local Similarity 100.0%; Pred. No. 4.5e-219;  
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 58 AISVTMDNIIISGNEPENVYILKPLAGVYRSLLKQIEKNITFNLDNLINSLKRRKY 117  
Db 1 AISVTMDNIIISGNEPENVYILKPLAGVYRSLLKQIEKNITFNLDNLINSLKRRKY 60  
QY 118 FLDVLESDLMQFHHISSNEVYIIEDSFKLLNSEQNTLLKSKYIKESVENDIKFAQEGIS 177  
Db 61 FLDVLESDLMQFHHISSNEVYIIEDSFKLLNSEQNTLLKSKYIKESVENDIKFAQEGIS 120  
QY 178 YYEKVLAKYKDDLESIRKKVKEKEKEFPSPPTPPSPAKTDEQKESKFLPFLTNIETL 237  
Db 121 YYEKVLAKYKDDLESIRKKVKEKEKEFPSPPTPPSPAKTDEQKESKFLPFLTNIETL 180  
QY 238 YNNLVNKIDDYILNKLAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFKN 288  
Db 181 YNNLVNKIDDYILNKLAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFKN 231

RESULT 2

US-08-195-705-4  
; Sequence 4, Application US/08195705  
; Patent No. 6420523  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Sandra  
; APPLICANT: Hul, George  
; APPLICANT: Barr, Philip  
; APPLICANT: Gibson, Helen  
; TITLE OF INVENTION: BACULOVIRUS PRODUCED PLASMIDIUM  
; TITLE OF INVENTION: FALCIPARUM VACCINE  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Davis Hoxie Faithfull Hapgood  
; STREET: 45 Rockefeller Pl.  
; CITY: New York  
; STATE: N.Y.  
; COUNTRY: USA  
; ZIP: 10111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/195,705  
; FILING DATE: 14-FEB-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jacobs, Seth H  
; REGISTRATION NUMBER: 32140  
; REFERENCE/DOCKET NUMBER: 11880A3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-757-2200  
; TELEFAX: 212-586-1461  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 394 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; FRAGMENT TYPE: C-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: Plasmodium falciparum

STRAIN: MAD  
US-08-195-705-4

Query Match 31.3%; Score 135; DB 4; Length 394;  
Best Local Similarity 100.0%; Pred. No. 1.2e-124; Indels 0; Gaps 0;  
Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 154 LKSYKIKESVNDIKFAQEGISYVEKVLAKYKDDLESIKKVIKEKEKFPSPPTTP 213  
DB 97 LKSYKIKESVNDIKFAQEGISYVEKVLAKYKDDLESIKKVIKEKEKFPSPPTTP 156  
  
QY 214 SPKATDEOKKESKFLPFLNIETLYNNLVNKIDYLLINKAKINDCNVKEDEAHVKITKL 273  
DB 157 SPKATDEOKKESKFLPFLNIETLYNNLVNKIDYLLINKAKINDCNVKEDEAHVKITKL 216  
  
QY 274 SLDKATDDKIDLFKN 288  
DB 217 SLDKATDDKIDLFKN 231

RESULT 3  
US-08-290-919-4  
; Sequence 4, Application US/08290919  
; Patent No. 5720959  
; GENERAL INFORMATION:  
; APPLICANT: HOLDER, ANTHONY A.  
; APPLICANT: BLACKMAN, MICHAEL J.  
; APPLICANT: CHAPPEL, JONATHAN A.  
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA  
; TITLE OF INVENTION: VACCINE  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.  
; STREET: 1100 NEW YORK AVENUE, N.W.  
; CITY: WASHINGTON  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-3918  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/290,919  
; FILING DATE: 04-OCT-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9203821.5  
; FILING DATE: 22-FEB-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/GB93/00367  
; FILING DATE: 22-FEB-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: KOKULIS, PAUL N.  
; REGISTRATION NUMBER: 16,773  
; REFERENCE/DOCKET NUMBER: 212242/HCM/MJL/6BC8/  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 861-3000  
; TELEFAX: (202) 822-0944  
; TELEX: 6714627 CUSH  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 53 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: 1  
; OTHER INFORMATION: /label=X  
; OTHER INFORMATION: /note="X = M and N, or N"

US-08-290-919-4

Query Match 11.6%; Score 50; DB 1; Length 53;  
Best Local Similarity 100.0%; Pred. No. 7.9e-42; Indels 0; Gaps 0;  
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 382 PNPTCNENGGCDADATCTEEDSGSSRKKITCCTCRPDSYPLFDGIFCSS 431  
DB 2 PNPTCNENGGCDADATCTEEDSGSSRKKITCCTCRPDSYPLFDGIFCSS 51

RESULT 4  
US-09-741-243C-4  
; Sequence 4, Application US/09741243C  
; Patent No. 6399352  
; GENERAL INFORMATION:  
; APPLICANT: Crawford Jr., John Milton  
; APPLICANT: Rice, John  
; APPLICANT: Sevala, Veeresh  
; APPLICANT: Stewart, Sandy  
; TITLE OF INVENTION: NOVEL PLANT PORPHOBILINOGEN AND FUSION  
; TITLE OF INVENTION: PROTEIN THEREOF  
; FILE REFERENCE: 2022US  
; CURRENT APPLICATION NUMBER: US/09/741,243C  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 60/171,785  
; PRIOR FILING DATE: 1999-12-22  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 167  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Thioedoxin functional fragment  
US-09-741-243C-4

Query Match 11.6%; Score 50; DB 4; Length 167;  
Best Local Similarity 100.0%; Pred. No. 2.2e-41; Indels 0; Gaps 0;  
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MHHHHHSSGLVPRGSGMKETAATAKFERQHMDSPDLGTDDDDKAMADIGS 50  
DB 116 MHHHHHSSGLVPRGSGMKETAATAKFERQHMDSPDLGTDDDDKAMADIGS 165

RESULT 5  
US-09-513-442-2  
; Sequence 2, Application US/09513442  
; Patent No. 6387664  
; GENERAL INFORMATION:  
; APPLICANT: Ikemoto, Mitsushi  
; TITLE OF INVENTION: SPARC Fusion Protein and a Process for Producing the  
; TITLE OF INVENTION: Same  
; FILE REFERENCE: HIRAKI-04218  
; CURRENT APPLICATION NUMBER: US/09/513,442  
; CURRENT FILING DATE: 2000-02-25  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 467  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-513-442-2

Query Match 11.6%; Score 50; DB 4; Length 467;  
Best Local Similarity 100.0%; Pred. No. 5.4e-41; Indels 0; Gaps 0;  
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MHHHHHSSGLVPRGSGMKETAATAKFERQHMDSPDLGTDDDDKAMADIGS 50  
DB 116 MHHHHHSSGLVPRGSGMKETAATAKFERQHMDSPDLGTDDDDKAMADIGS 165

## RESULT 6

US-09-741-243C-2  
; Sequence 2, Application US/09741243C  
; Patent No. 6399352  
; GENERAL INFORMATION:  
; APPLICANT: Crawford Jr., John Milton  
; APPLICANT: Rice, John  
; APPLICANT: Sevala, Veeresh  
; APPLICANT: Stewart, Sandy  
; TITLE OF INVENTION: NOVEL PLANT PORPHOBILINOGEN AND FUSION  
; TITLE OF INVENTION: PROTEIN THEREOF  
; FILE REFERENCE: 2022US  
; CURRENT APPLICATION NUMBER: US/09/741.243C  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 60/171,785  
; PRIOR FILING DATE: 1999-12-22  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 551  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: A plant thioredoxin-porphobilinogen  
; OTHER INFORMATION: synthase fusion protein  
US-09-741-243C-2

Query Match 11.6%; Score 50; DB 4; Length 551;  
Best Local Similarity 100.0%; Pred. No. 6.3e-41;  
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHHHHHSSGLVPRGSGMKETAAAFERQHMDSPDLGTDGDDDKAMADIGS 50  
|||||  
Db 116 MHHHHHSSGLVPRGSGMKETAAAFERQHMDSPDLGTDGDDDKAMADIGS 165  
|||||

## RESULT 7

US-09-626-589-3  
; Sequence 3, Application US/09626589  
; Patent No. 6326164  
; GENERAL INFORMATION:  
; APPLICANT: Rice, John  
; APPLICANT: Kloti, Andreas  
; APPLICANT: Crawford, John  
; APPLICANT: Lanning, Beth  
; APPLICANT: Stewart, Sandy  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE  
; TITLE OF INVENTION: IDENTIFICATION OF MODULATORS OF DEOXYXYLULOSE 5-PHOSPHATE  
; TITLE OF INVENTION: SYNTHASE ACTIVITY  
; FILE REFERENCE: 2037 US  
; CURRENT APPLICATION NUMBER: US/09/626.589  
; CURRENT FILING DATE: 2000-07-27  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 824  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Residues 1-165 are from the thioredoxin sequence  
; OTHER INFORMATION: found in the vector pET32 supplied by No. 6326164agen.  
; OTHER INFORMATION: Residues 166-824 represent the tDXPS sequence from  
; OTHER INFORMATION: Arabidopsis shown in SEQ ID NO:2.  
US-09-626-589-3

Query Match 11.6%; Score 50; DB 4; Length 824;  
Best Local Similarity 100.0%; Pred. No. 8.9e-41;  
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHHHHHSSGLVPRGSGMKETAAAFERQHMDSPDLGTDGDDDKAMADIGS 50  
|||||  
Db 116 MHHHHHSSGLVPRGSGMKETAAAFERQHMDSPDLGTDGDDDKAMADIGS 165  
|||||

## RESULT 8

US-08-290-919-2  
; Sequence 2, Application US/08290919  
; Patent No. 5720959  
; GENERAL INFORMATION:  
; APPLICANT: HOLDER, ANTHONY A.  
; APPLICANT: BLACKMAN, MICHAEL J.  
; APPLICANT: CHAPPEL, JONATHAN A.  
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA  
; TITLE OF INVENTION: VACCINE  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.  
; STREET: 1100 NEW YORK AVENUE, N.W.  
; CITY: WASHINGTON  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-3918  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/290.919  
; FILING DATE: 04-OCT-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9203821.5  
; FILING DATE: 22-FEB-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/GB93/00367  
; FILING DATE: 22-FEB-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: KOKULIS, PAUL N.  
; REGISTRATION NUMBER: 16,773  
; REFERENCE/DOCKET NUMBER: 212242/HCM/MJL/6BC8/  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 861-3000  
; TELEFAX: (202) 822-0944  
; TELEX: 6714627 CUSH  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 48 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: 1  
; OTHER INFORMATION: /label= X  
; OTHER INFORMATION: /note= "X = M and N, or N"  
US-08-290-919-2

Query Match 10.9%; Score 47; DB 1; Length 48;  
Best Local Similarity 100.0%; Pred. No. 6.4e-39;  
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 339 ISQHCVKQCPENSGCFRHLDEREECKLLNKKQEGDKCVENPNT 385  
|||||  
Db 2 ISQHCVKQCPENSGCFRHLDEREECKLLNKKQEGDKCVENPNT 48  
|||||

## RESULT 9

US-08-683-007A-2  
; Sequence 2, Application US/08683007A  
; Patent No. 5858724  
; GENERAL INFORMATION:  
; APPLICANT: No. 5858724y, Robert E  
; APPLICANT: Domanico, Michael

APPLICANT: Yaeger, Keith  
APPLICANT: Kroeker, Warren  
TITLE OF INVENTION: Recombinant Rabbit Tissue Factor  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Quarles & Brady  
STREET: 1 South Pinckney Street  
CITY: Madison  
STATE: WI  
COUNTRY: USA  
ZIP: 53703  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/683,007A  
FILING DATE: 16-JUL-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Seay, Nicholas J  
REGISTRATION NUMBER: 27386  
REFERENCE/DOCKET NUMBER: 740380.90040  
TELEPHONE: 608-251-5000  
TELEFAX: 608-251-9166  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 408 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-683-007A-2

Query Match 10.7%; Score 46; DB 2; Length 408;  
Best Local Similarity 100.0%; Pred. No. 4.1e-37;  
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MHHHHSSGLVPRGSMKETAARKFERQHMDSPDLGTDDDDKAMA 46  
|||||  
DB 116 MHHHHSSGLVPRGSMKETAARKFERQHMDSPDLGTDDDDKAMA 161

RESULT 10  
US-08-195-705-5  
Sequence 5, Application US/08195705  
Patent No. 6420523  
GENERAL INFORMATION:  
APPLICANT: Chang, Sandra  
APPLICANT: Hui, George  
APPLICANT: Barr, Philip  
APPLICANT: Gibson, Helen  
TITLE OF INVENTION: BACULOVIRUS PRODUCED PLASMODIUM  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Davis Hoxie Faithfull Hapgood  
STREET: 45 Rockefeller Pl.  
CITY: New York  
STATE: N.Y.  
COUNTRY: USA  
ZIP: 10111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/195,705  
FILING DATE: 14-FEB-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:

NAME: Jacobs, Seth H  
REGISTRATION NUMBER: 32140  
REFERENCE/DOCKET NUMBER: 11880A3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-757-2200  
TELEFAX: 212-586-1461  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 377 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
FRAGMENT TYPE: C-terminal  
ORIGINAL SOURCE:  
ORGANISM: Plasmodium falciparum  
STRAIN: WEL  
US-08-195-705-5  
Query Match 10.4%; Score 45; DB 4; Length 377;  
Best Local Similarity 100.0%; Pred. No. 3.7e-36;  
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 353 SGCFRHLDERECKLLNLYKQEGDKCVENPNTCNENNGCDADA 397  
|||||  
DB 279 SGCFRHLDERECKLLNLYKQEGDKCVENPNTCNENNGCDADA 323  
RESULT 11  
US-08-290-919-1  
Sequence 1, Application US/08290919  
Patent No. 5720959  
GENERAL INFORMATION:  
APPLICANT: HOLDER, ANTHONY A.  
APPLICANT: BLACKMAN, MICHAEL J.  
APPLICANT: CHAPPEL, JONATHAN A.  
TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.  
STREET: 1100 NEW YORK AVENUE, N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005-3918  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/290,919  
FILING DATE: 04-OCT-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9203821.5  
FILING DATE: 22-FEB-1992  
PRIOR APPLICATION DATA: PCT/GB93/00367  
FILING DATE: 22-FEB-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: KOKULIS, PAUL N.  
REGISTRATION NUMBER: 16,773  
REFERENCE/DOCKET NUMBER: 212242/HCM/NJL/6BC8/  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 861-3000  
TELEFAX: (202) 822-0944  
TELEX: 6714627 CUSH  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 48 amino acids

TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1  
OTHER INFORMATION: /label- x  
OTHER INFORMATION: /note= "x" = M and N, or N"  
US-08-290-919-1  
Query Match 7.9%; Score 34; DB 1; Length 48;  
Best Local Similarity 100.0%; Pred. No. 3.9e-26;  
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 352 NSGCFRHLDERECKCLLNKQEGDKCVENPNT 385  
Db 15 NSGCFRHLDERECKCLLNKQEGDKCVENPNT 48  
RESULT 12  
US-08-290-919-12  
Sequence 12, Application US/08290919  
Patent No. 5720959  
GENERAL INFORMATION:  
APPLICANT: HOLDER, ANTHONY A.  
APPLICANT: BLACKMAN, MICHAEL J.  
APPLICANT: CHAPPEL, JONATHAN A.  
TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA  
TITLE OF INVENTION: VACCINE  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CUSHMAN DABY & CUSHMAN, L.L.P.  
STREET: 1100 NEW YORK AVENUE, N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005-3918  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/290,919  
FILING DATE: 04-OCT-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9203821.5  
FILING DATE: 22-FEB-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB93/00367  
FILING DATE: 22-FEB-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: KOKULIS, PAUL N.  
REGISTRATION NUMBER: 16,773  
REFERENCE/DOCKET NUMBER: 212242/HCM/MJL/6BC8/  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 861-3000  
TELEFAX: (202) 822-0944  
TELEX: 6714627 CUSH  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 48 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-290-919-12  
Query Match 7.9%; Score 34; DB 1; Length 48;  
Best Local Similarity 100.0%; Pred. No. 3.9e-26;  
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 NSGCFRHLDERECKCLLNKQEGDKCVENPNT 385  
Db 15 NSGCFRHLDERECKCLLNKQEGDKCVENPNT 48  
RESULT 13  
US-09-166-966E-11  
Sequence 11, Application US/09166966E  
Patent No. 6407208  
GENERAL INFORMATION:  
APPLICANT: CHEN, DAVID CHANHAN  
APPLICANT: HU, NIEN-TAI  
APPLICANT: CHEN, YUN-JU  
APPLICANT: HSEU, TZONG-HSIUNG  
TITLE OF INVENTION: CHIMERIC PROTEINS WITH A CELLULOSE BINDING DOMAIN  
FILE REFERENCE: 32350-150960  
CURRENT APPLICATION NUMBER: US/09/166,966E  
CURRENT FILING DATE: 1998-10-06  
PRIOR APPLICATION NUMBER: TW 86114750  
PRIOR FILING DATE: 1997-10-08  
NUMBER OF SEQ ID NOS: 11  
SEQ ID NO 11  
LENGTH: 209  
TYPE: PRT  
ORGANISM: Unknown  
FEATURE:  
NAME/KEY: thioredoxin  
LOCATION: 1..109  
NAME/KEY: Cellulose-binding domain  
LOCATION: 162..197  
NAME/KEY: RGD  
LOCATION: 204..206  
OTHER INFORMATION: Combined amino acid sequence of thioredoxin,  
OTHER INFORMATION: artificial sequence, cellulose-binding domain, and  
OTHER INFORMATION: artificial sequence; the two artificial sequences  
OTHER INFORMATION: flanking the cellulose-binding domain contain restriction site  
US-09-166-966E-11  
Query Match 6.3%; Score 27; DB 4; Length 209;  
Best Local Similarity 100.0%; Pred. No. 1.1e-18;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 ETAAAKFERQHMDSPDLGTDGDDKAMA 46  
Db 135 ETAAAKFERQHMDSPDLGTDGDDKAMA 161  
RESULT 14  
US-09-277-716-22  
Sequence 22, Application US/09277716A  
Patent No. 6232107  
GENERAL INFORMATION:  
APPLICANT: Bryan, Bruce  
APPLICANT: Szent-Gyorgyi, Christopher  
APPLICANT: PROLUME, LTD.  
TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE  
CURRENT APPLICATION NUMBER: US/09/277,716A  
CURRENT FILING DATE: 1999-03-26  
EARLIER APPLICATION NUMBER: 60/102,939  
EARLIER FILING DATE: 1998-10-01  
EARLIER APPLICATION NUMBER: 60/089,367  
EARLIER FILING DATE: 1998-06-15  
EARLIER APPLICATION NUMBER: 60/079,624  
EARLIER FILING DATE: 1998-03-27  
NUMBER OF SEQ ID NOS: 32  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 22  
LENGTH: 382  
TYPE: PRT  
ORGANISM: Artificial Sequence: fusion protein  
FEATURE:  
OTHER INFORMATION: Cellulose Binding Domain (CBD)-Gaussia<sup>1</sup> luciferase fusion prote

US-09-277-716-22

Query Match 6.3%; Score 27; DB 4; Length 382;  
Best Local Similarity 100.0%; Pred. No. 1.9e-18;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 GKKETAAAKFERQHMDSPDLGTDDDK 43  
DB 171 GKKETAAAKFERQHMDSPDLGTDDDK 197

RESULT 15

US-09-609-161B-22

; Sequence 22, Application US/09609161B

; Patent No. 6436682

; GENERAL INFORMATION:

; APPLICANT: Bryan, Bruce

; APPLICANT: Szent-Gyorgyi, Christopher

; APPLICANT: PROLUME, LTD.

; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE LUC

; TITLE OF INVENTION: AND FLUORESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HI

; TITLE OF INVENTION: SCREENING AND NOVELTY ITEMS

; FILE REFERENCE: 24729-121B

; CURRENT APPLICATION NUMBER: US/09/609,161B

; CURRENT FILING DATE: 2000-06-30

; PRIOR APPLICATION NUMBER: 09/277,716

; PRIOR FILING DATE: 1999-03-26

; PRIOR APPLICATION NUMBER: 60/102,939

; PRIOR FILING DATE: 1998-10-01

; PRIOR APPLICATION NUMBER: 60/089,367

; PRIOR FILING DATE: 1998-06-15

; PRIOR APPLICATION NUMBER: 60/079,624

; PRIOR FILING DATE: 1998-03-27

; NUMBER OF SEQ ID NOS: 32

; SOFTWARE: Patent In Ver. 2.0

; SEQ ID NO 22

; LENGTH: 382

; TYPE: PRT

; ORGANISM: Artificial Sequence: fusion protein

; FEATURE:

; OTHER INFORMATION: Cellulose Binding Domain (CBD)-Gaussia luciferase fusion protein

US-09-609-161B-22

Query Match 6.3%; Score 27; DB 4; Length 382;

Best Local Similarity 100.0%; Pred. No. 1.9e-18;

Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 GKKETAAAKFERQHMDSPDLGTDDDK 43

DB 171 GKKETAAAKFERQHMDSPDLGTDDDK 197

RESULT 16

US-09-166-966E-8

; Sequence 8, Application US/09166966E

; Patent No. 6407208

; GENERAL INFORMATION:

; APPLICANT: CHEN, DAVID CHANHAN

; APPLICANT: HU, NIEN-TAI

; APPLICANT: CHEN, YUN-JU

; APPLICANT: HSEU, TZONG-HSIUNG

; TITLE OF INVENTION: CHIMERIC PROTEINS WITH A CELLULOSE BINDING DOMAIN

; FILE REFERENCE: 32350-150960

; CURRENT APPLICATION NUMBER: US/09/166,966E

; CURRENT FILING DATE: 1998-10-06

; PRIOR APPLICATION NUMBER: TW 86114750

; PRIOR FILING DATE: 1997-10-08

; NUMBER OF SEQ ID NOS: 11

; SEQ ID NO 8

; LENGTH: 159

; TYPE: PRT

; ORGANISM: Unknown

; FEATURE:

; NAME/KEY: thioredoxin

; LOCATION: 1..109

; OTHER INFORMATION: Combined amino acid of thioredoxin and an artificial sequence

US-09-166-966E-8

Query Match 5.6%; Score 24; DB 4; Length 159;

Best Local Similarity 100.0%; Pred. No. 7.7e-16;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 ETAAAKFERQHMDSPDLGTDDDK 43

DB 135 ETAAAKFERQHMDSPDLGTDDDK 158

RESULT 17

US-08-195-705-3

; Sequence 3, Application US/08195705

; Patent No. 6420523

; GENERAL INFORMATION:

; APPLICANT: Chang, Sandra

; APPLICANT: Hui, George

; APPLICANT: Barr, Philip

; APPLICANT: Gibson, Helen

; TITLE OF INVENTION: BACULOVIRUS PRODUCED PLASMODIUM

; TITLE OF INVENTION: FALCIPARUM VACCINE

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Davis Hoxie Faithfull Hapgood

; STREET: 45 Rockefeller Pl.

; CITY: New York

; STATE: N.Y.

; COUNTRY: USA

; ZIP: 10111

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/195,705

; FILING DATE: 14-FEB-1994

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Jacobs, Seth H

; REGISTRATION NUMBER: 32140

; REFERENCE/DOCKET NUMBER: 11880A3

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-757-2200

; TELEFAX: 212-586-1461

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 375 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; HYPOTHETICAL: NO

; FRAGMENT TYPE: C-terminal

; ORIGINAL SOURCE:

; ORGANISM: Plasmodium falciparum

; STRAIN: KI

US-08-195-705-3

Query Match 5.6%; Score 24; DB 4; Length 375;

Best Local Similarity 100.0%; Pred. No. 1.7e-15;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 NSGCFRHLDERECKCLLNKQEG 375

DB 276 NSGCFRHLDERECKCLLNKQEG 299

RESULT 18

```

: TITLE OF INVENTION: Transforming Growth Factor B Fusion
: TITLE OF INVENTION: and
: TITLE OF INVENTION: Their Use in Wound Healing
: NUMBER OF SEQUENCES: 34
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Merchant & Gould
: STREET: 11150 Santa Monica Boulevard, Suite 400
: CITY: Los Angeles
: STATE: California
: COUNTRY: USA
: ZIP: 90025-3395
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/470,837
: FILING DATE:
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: Sharp, Janice A.
: REGISTRATION NUMBER: 34,051
: REFERENCE/DOCKET NUMBER: 30630-IUS01
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 310-445-1140
: TELEFAX: 310-445-9031
: INFORMATION FOR SEQ ID NO: 24:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 15 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-470-837-24

Query Match 3.5%; Score 15; DB 1; Length 15:1
Best Local Similarity 100.0%; Pred. No. 6.8e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0

QY 19 KETAAAKFERQHMD5 33
Db 1 KETAAAKFERQHMD5 15

RESULT 20
US-08-639-806-1
: Sequence 1, Application US/08639806
: Patent No. 5817455
: GENERAL INFORMATION:
: APPLICANT: Raines, Ronald T
: TITLE OF INVENTION: Method for Enzyme Inactivation
: NUMBER OF SEQUENCES: 1
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Quarles and Brady
: STREET: 1 South Pinckney Street
: CITY: Madison
: STATE: WI
: COUNTRY: USA
: ZIP: 53703
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/639,806
: FILING DATE: 29-APR-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/203,536
: FILING DATE: 01-MAR-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Seay, Nicholas J

```



REGISTRATION NUMBER: 27,386  
REFERENCE/DOCKET NUMBER: 70-399-9002-9  
TELEPHONE: 608-251-5000  
TELEFAX: 608-251-9166  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
US-08-639-806-1

Query Match 3.5%; Score 15; DB 2; Length 15;  
Best Local Similarity 100.0%; Pred. No. 6.8e-08;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KETAAKFERQHMD 33  
|||||

DB 1 KETAAKFERQHMD 15

## RESULT 21

US-08-868-452-24  
Sequence 24, Application US/08868452C  
Patent No. 6352972

GENERAL INFORMATION:  
APPLICANT: Marcel E. Nimni

APPLICANT: Frederick L. Hall

APPLICANT: Lingtao Wu

APPLICANT: Bo Han

APPLICANT: Edwin Shors

TITLE OF INVENTION: BONE MORPHOGENETIC PROTEINS AND THEIR

TITLE OF INVENTION: USE IN BONE GROWTH

FILE REFERENCE: 17972-11

CURRENT APPLICATION NUMBER: US/08/868,452C

CURRENT FILING DATE: 1997-06-03

NUMBER OF SEQ ID NOS: 51

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 24

LENGTH: 15

TYPE: PRT

ORGANISM: Human

US-08-868-452-24

Query Match 3.5%; Score 15; DB 4; Length 15;  
Best Local Similarity 100.0%; Pred. No. 6.8e-08;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KETAAKFERQHMD 33  
|||||

DB 1 KETAAKFERQHMD 15

## RESULT 22

US-08-373-134D-3

Sequence 3, Application US/08373134D

Patent No. 5780296

GENERAL INFORMATION:

APPLICANT: Kmiec, Eric

APPLICANT: Holloman, William

TITLE OF INVENTION: COMPOSITIONS AND METHODS TO PROMOTE

TITLE OF INVENTION: HOMOLOGOUS RECOMBINATION IN EUKARYOTIC CELLS AND ORGANISMS

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/373,134D  
FILING DATE: January 17, 1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Friebe, Thomas E.

REGISTRATION NUMBER: 29,258

REFERENCE/DOCKET NUMBER: 7991-007

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-9741/8864

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 20 amino acids

TYPE: amino acid

TOPOLOGY: unknown

MOLECULE TYPE: peptide

US-08-373-134D-3

Query Match 3.5%; Score 15; DB 1; Length 20;  
Best Local Similarity 100.0%; Pred. No. 8.8e-08;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HHHHHSSGLVPRGS 16

|||||

DB 5 HHHHHSSGLVPRGS 19

## RESULT 23

US-09-114-637-3

Sequence 3, Application US/09114637

Patent No. 5945339

GENERAL INFORMATION:

APPLICANT: Kmiec, Eric

APPLICANT: Holloman, William

TITLE OF INVENTION: COMPOSITIONS AND METHODS TO PROMOTE

TITLE OF INVENTION: HOMOLOGOUS RECOMBINATION IN EUKARYOTIC CELLS AND ORGANISMS

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/114,637

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/373,134

FILING DATE: January 17, 1995

ATTORNEY/AGENT INFORMATION:

NAME: Friebe, Thomas E.

REGISTRATION NUMBER: 29,258

REFERENCE/DOCKET NUMBER: 7991-007

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-9741/8864

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 20 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
US-09-114-637-3

Query Match 3.5%; Score 15; DB 2; Length 20;  
Best Local Similarity 100.0%; Pred. No. 8.8e-08;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HHHHHSSGLVPRGS 16  
| | | | | | | | | | | | | | | | | | | | | |  
Db 5 HHHHHSSGLVPRGS 19

## RESULT 24

US-08-485-942A-99  
; Sequence 99, Application US/08485942A  
; Patent No. 6048837  
; GENERAL INFORMATION:  
; APPLICANT: JEFFREY M. FRIEDMAN, YIYING ZHANG, RICARDO PROENCA,  
; APPLICANT: MARGHERITA MAFFEI, JEFFREY HALAAS, KETAN GAJIWALA, AND STEPHEN K. BURLE  
; TITLE OF INVENTION: OB POLYPEPTIDE AS MODULATORS OF BODY WEIGHT (AS  
; TITLE OF INVENTION: AMENDED)  
; NUMBER OF SEQUENCES: 99  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Klauber & Jackson  
; STREET: 411 Hackensack Avenue  
; CITY: Hackensack  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07601

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/485.942A  
; FILING DATE: JUNE 7, 1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/438,431  
; FILING DATE: May 10, 1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/347,563  
; FILING DATE: June 30, 1994  
; CLASSIFICATION:  
; APPLICATION DATA:  
; APPLICATION NUMBER: 08/292,345  
; FILING DATE: August 17, 1994  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jackson Esq., David A.  
; REGISTRATION NUMBER: 26,742  
; REFERENCE/DOCKET NUMBER: 600-1-087 CIP 2F  
; TELEPHONE: 201 487-5800  
; TELEFAX: 201 343-1684  
; TELEX: 133521

; INFORMATION FOR SEQ ID NO: 99:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 20 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FRAGMENT TYPE: N-terminal His-tag  
US-08-485-942A-99

Query Match 3.5%; Score 15; DB 3; Length 20;  
Best Local Similarity 100.0%; Pred. No. 8.8e-08;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 2 HHHHHSSGLVPRGS 16  
| | | | | | | | | | | | | | | | | | | | | |  
Db 5 HHHHHSSGLVPRGS 19

## RESULT 25

US-08-488-214A-99  
; Sequence 99, Application US/08488214A  
; Patent No. 6124439  
; GENERAL INFORMATION:  
; APPLICANT: JEFFREY M. FRIEDMAN, YIYING ZHANG, RICARDO PROENCA,  
; APPLICANT: MARGHERITA MAFFEI, JEFFREY HALAAS, KETAN GAJIWALA, AND STEPHEN K. BU  
; TITLE OF INVENTION: OB POLYPEPTIDE ANTIBODIES AND METHOD OF MAKING  
; TITLE OF INVENTION: (AS AMENDED)  
; NUMBER OF SEQUENCES: 99  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Klauber & Jackson  
; STREET: 411 Hackensack Avenue  
; CITY: Hackensack  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07601

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/488.214A  
; FILING DATE: JUNE 7, 1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/438,431  
; FILING DATE: May 10, 1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/347,563  
; FILING DATE: No. 6124439ember 30, 1994  
; CLASSIFICATION:  
; APPLICATION DATA:  
; APPLICATION NUMBER: 08/292,345  
; FILING DATE: August 17, 1994  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jackson Esq., David A.  
; REGISTRATION NUMBER: 26,742  
; REFERENCE/DOCKET NUMBER: 600-1-087 CIP 2D  
; TELEPHONE: 201 487-5800  
; TELEFAX: 201 343-1684  
; TELEX: 133521

; INFORMATION FOR SEQ ID NO: 99:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 20 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FRAGMENT TYPE: N-terminal His-tag  
US-08-488-214A-99

Query Match 3.5%; Score 15; DB 3; Length 20;  
Best Local Similarity 100.0%; Pred. No. 8.8e-08;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HHHHHSSGLVPRGS 16  
| | | | | | | | | | | | | | | | | | | | | |  
Db 5 HHHHHSSGLVPRGS 19

## RESULT 26

US-08-438-431A-99

```
; Sequence 99, Application US/08438431A
; Patent No. 6429290
; GENERAL INFORMATION:
; APPLICANT: JEFFREY M. FRIEDMAN, YIYING ZHANG, RICARDO PROENCA, MARGHERITA MAFFEI,
; TITLE OF INVENTION: MODULATORS OF BODY WEIGHT, CORRESPONDING NUCLEIC ACIDS AND PR
; NUMBER OF SEQUENCES: 99
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/438,431A
; FILING DATE: May 10, 1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/347,563
; FILING DATE: No. 6429290ember 30, 1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/292,345
; FILING DATE: August 17, 1994
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-087 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 99:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: N-terminal His-tag
; US-08-438-431A-99

Query Match 3.5%; Score 15; DB 4; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.8e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HHHHHSSGLVPRGS 16
Db 5 HHHHHSSGLVPRGS 19

; RESULT 27
; US-09-638-202A-114
; Sequence 114, Application US/09638202A
; Patent No. 6462189
; GENERAL INFORMATION:
; APPLICANT: Koteda, Shohel
; TITLE OF INVENTION: ARTIFICIAL ANTIBODY POLYPEPTIDES
; NUMBER OF SEQUENCES: 118
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth P.A.
; STREET: 121 South Eighth Street, Ste. 1600
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
```

```
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/638,202A
; FILING DATE: 11-Aug-2000
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/096,749
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ann S. Vilksins
; REGISTRATION NUMBER: 37,748
; REFERENCE/DOCKET NUMBER: 109.034US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (612) 373-6900
; TELEFAX: (612) 339-3061
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; SEQUENCE DESCRIPTION: SEQ ID NO: 114:
; US-09-638-202A-114

Query Match 3.5%; Score 15; DB 4; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.8e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HHHHHSSGLVPRGS 16
Db 5 HHHHHSSGLVPRGS 19

; RESULT 28
; US-08-432-871C-46
; Sequence 46, Application US/08432871C
; Patent No. 5877010
; GENERAL INFORMATION:
; APPLICANT: Loeb, Lawrence A.
; APPLICANT: Black, Margaret E.
; TITLE OF INVENTION: THYMIDINE KINASE MUTANTS
; NUMBER OF SEQUENCES: 104
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/432,871C
; FILING DATE: 02-MAY-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Mcmasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 240052.409C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836
; INFORMATION FOR SEQ ID NO: 46:
```

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 21 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-432-871C-46

Query Match 3.5%; Score 15; DB 2; Length 21;  
Best Local Similarity 100.0%; Pred. No. 9.2e-08;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 2 HHHHHSSGLVPRGS 16  
Db 5 HHHHHSSGLVPRGS 19

## RESULT 29

US-08-712-878-6  
; Sequence 6, Application US/08712878  
; Patent No. 5985863  
; GENERAL INFORMATION:  
; APPLICANT: Su, Michael  
; APPLICANT: Gu, Yong  
; APPLICANT: Livingston, David J.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DECREASING  
; TITLE OF INVENTION: IGIF AND IFN-GAMMA PRODUCTION BY ADMINISTERING AN ICE  
; TITLE OF INVENTION: INHIBITOR  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: James F. Haley, Jr.  
; STREET: 1251 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: United States of America  
; ZIP: 10020  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/712,878  
; FILING DATE: 12-SEP-1996  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haley Jr., James F.  
; REGISTRATION NUMBER: 27,794  
; REFERENCE/DOCKET NUMBER: VPI/96-05  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-596-9000  
; TELEFAX: 212-596-9090  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 21 amino acids  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-712-878-6

Query Match 3.5%; Score 15; DB 2; Length 21;  
Best Local Similarity 100.0%; Pred. No. 9.2e-08;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HHHHHSSGLVPRGS 16  
Db 5 HHHHHSSGLVPRGS 19

## RESULT 30

US-08-485-942A-98

; Sequence 98, Application US/08485942A  
; Patent No. 6048837  
; GENERAL INFORMATION:  
; APPLICANT: JEFFREY M. FRIEDMAN, YIYING ZHANG, RICARDO PROENCA,  
; APPLICANT: MARGHERITA MAFFEI, JEFFREY HALAAS, KETAN GAJIWALA, AND STEPHEN K. BU  
; TITLE OF INVENTION: OB POLYPEPTIDE AS MODULATORS OF BODY WEIGHT (AS  
; TITLE OF INVENTION: AMENDED)  
; NUMBER OF SEQUENCES: 99  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Klauber & Jackson  
; STREET: 411 Hackensack Avenue  
; CITY: Hackensack  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/485,942A  
; FILING DATE: JUNE 7, 1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/438,431  
; FILING DATE: May 10, 1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/347,563  
; FILING DATE: No. 6048837ember 30, 1994  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/292,345  
; FILING DATE: August 17, 1994  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jackson Esq., David A.  
; REGISTRATION NUMBER: 26,742  
; REFERENCE/DOCKET NUMBER: 600-1-087 CIP 2F  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201 487-5800  
; TELEFAX: 201 343-1684  
; TELEX: 133521  
; INFORMATION FOR SEQ ID NO: 98:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 21 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FRAGMENT TYPE: N-terminal  
US-08-485-942A-98

Query Match 3.5%; Score 15; DB 3; Length 21;  
Best Local Similarity 100.0%; Pred. No. 9.2e-08;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HHHHHSSGLVPRGS 16  
Db 5 HHHHHSSGLVPRGS 19

## RESULT 31

US-08-488-214A-98  
; Sequence 98, Application US/08488214A  
; Patent No. 6124439

; GENERAL INFORMATION:  
; APPLICANT: JEFFREY M. FRIEDMAN, YIYING ZHANG, RICARDO PROENCA,  
; APPLICANT: MARGHERITA MAFFEI, JEFFREY HALAAS, KETAN GAJIWALA, AND STEPHEN K. BU  
; TITLE OF INVENTION: OB POLYPEPTIDE ANTIBODIES AND METHOD OF MAKING  
; TITLE OF INVENTION: (AS AMENDED)  
; NUMBER OF SEQUENCES: 99  
; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: Klauber & Jackson  
;; STREET: 411 Hackensack Avenue  
;; CITY: Hackensack  
;; STATE: New Jersey  
;; COUNTRY: USA  
;; ZIP: 07601  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;;  
;; CURRENT APPLICATION DATA:  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; FILING DATE: June 7, 1995  
;; APPLICATION NUMBER: US/08/488,214A  
;;  
;; CLASSIFICATION:  
;; FILING DATE: June 7, 1995  
;;  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/438,431  
;; FILING DATE: May 10, 1995  
;; CLASSIFICATION:  
;; FILING DATE: August 17, 1994  
;;  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/347,563  
;; FILING DATE: NO. 6124439ember 30, 1994  
;; CLASSIFICATION:  
;; FILING DATE: May 10, 1995  
;;  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/292,345  
;; FILING DATE: August 17, 1994  
;; CLASSIFICATION:  
;; FILING DATE: August 17, 1994  
;;  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Jackson Esq., David A.  
;; REGISTRATION NUMBER: 26,742  
;; REFERENCE/DOCKET NUMBER: 600-1-087 CIP 2D  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 201 487-5800  
;; TELEFAX: 201 343-1684  
;; TELEX: 133521  
;;  
;; INFORMATION FOR SEQ ID NO: 98:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 21 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; FRAGMENT TYPE: N-terminal  
;;  
;; US-08-488-214A-98

Query Match 3.5%; Score 15; DB 3; Length 21;  
Best Local Similarity 100.0%; Pred. No. 9.2e-08;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HHHHHSSGLVPRGS 16  
Db 5 HHHHHSSGLVPRGS 19

RESULT 32  
US-08-488-208A-98  
; Sequence 98, Application US/08488208A  
; Patent No. 612448  
; GENERAL INFORMATION:  
; APPLICANT: THE ROCKEFELLER UNIVERSITY  
; TITLE OF INVENTION: MODULATORS OF BODY WEIGHT, CORRESPONDING  
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS, AND DIAGNOSTIC AND THERAPEUTIC  
; TITLE OF INVENTION: USES THEREOF  
; NUMBER OF SEQUENCES: 98  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Klauber & Jackson  
; STREET: 411 Hackensack Avenue  
; CITY: Hackensack  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; FILING DATE: 07-JUN-1995  
; APPLICATION NUMBER: US/08/483,211A  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:

;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/488,208A  
;; FILING DATE: 07-JUN-1995  
;; CLASSIFICATION: 514  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/485,943  
;; FILING DATE: June 7, 1995  
;; APPLICATION NUMBER: 08/438,431  
;; FILING DATE: May 10, 1995  
;; CLASSIFICATION: 514  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/347,563  
;; FILING DATE: NO. 6124448ember 30, 1994  
;; CLASSIFICATION: 514  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/292,345  
;; FILING DATE: August 17, 1994  
;; CLASSIFICATION: 514  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Jackson Esq., David A.  
;; REGISTRATION NUMBER: 26,742  
;; REFERENCE/DOCKET NUMBER: 600-1-087 CIP2I  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 201 487-5800  
;; TELEFAX: 201 343-1684  
;; TELEX: 133521  
;;  
;; INFORMATION FOR SEQ ID NO: 98:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 21 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; FRAGMENT TYPE: N-terminal  
;;  
;; US-08-488-208A-98

Query Match 3.5%; Score 15; DB 3; Length 21;  
Best Local Similarity 100.0%; Pred. No. 9.2e-08;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HHHHHSSGLVPRGS 16  
Db 5 HHHHHSSGLVPRGS 19

RESULT 33  
US-08-483-211A-98  
; Sequence 98, Application US/08483211A  
; Patent No. 6309853  
; GENERAL INFORMATION:  
; APPLICANT: THE ROCKEFELLER UNIVERSITY  
; TITLE OF INVENTION: MODULATORS OF BODY WEIGHT, CORRESPONDING  
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS, AND DIAGNOSTIC AND THERAPEUTIC  
; TITLE OF INVENTION: USES THEREOF  
; NUMBER OF SEQUENCES: 98  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Klauber & Jackson  
; STREET: 411 Hackensack Avenue  
; CITY: Hackensack  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; FILING DATE: 07-JUN-1995  
; APPLICATION NUMBER: US/08/483,211A  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/485,943  
FILING DATE: June 7, 1995  
APPLICATION NUMBER: 08/438,431  
FILING DATE: May 10, 1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/347,563  
FILING DATE: No. 630985ember 30, 1994  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/292,345  
FILING DATE: August 17, 1994  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-087 CIP21  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 98:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-483-211A-98

Query Match 3.5% Score 15; DB 4; Length 21;  
Best Local Similarity 100.0%; Pred. No. 9.2e-08;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HHHHHSSGLVPRGS 16  
| | | | | | | | | | | | | | | | | | | | | |  
Db 5 HHHHHSSGLVPRGS 19

RESULT 34  
US-08-488-223A-98

Sequence 98, Application US/08488223A  
Patent No. 6350730

## GENERAL INFORMATION:

TITLE OF INVENTION: THE ROCKEFELLER UNIVERSITY  
ACIDS AND PROTEINS, AND DIAGNOSTIC AND THERAPEUTIC USES THE

NUMBER OF SEQUENCES: 98

CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauber & Jackson

STREET: 411 Hackensack Avenue

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/488,223A

FILING DATE: 07-Jun-1995

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/485,943

FILING DATE: <Unknown>

APPLICATION NUMBER: 08/347,563

FILING DATE: No. 6350730ember 30, 1994

APPLICATION NUMBER: 08/292,345

FILING DATE: August 17, 1994

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-087 CIP21  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 98:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
SEQUENCE DESCRIPTION: SEQ ID NO: 98:  
US-08-488-223A-98

Query Match 3.5% Score 15; DB 4; Length 21;  
Best Local Similarity 100.0%; Pred. No. 9.2e-08;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HHHHHSSGLVPRGS 16  
| | | | | | | | | | | | | | | | | | | | | |  
Db 5 HHHHHSSGLVPRGS 19

RESULT 35

US-08-438-431A-98

Sequence 98, Application US/08438431A

Patent No. 6429290

GENERAL INFORMATION:

TITLE OF INVENTION: MODULATORS OF BODY WEIGHT, RICARDO PROENCA, MARGHERITA MAFEE  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauber & Jackson

STREET: 411 Hackensack Avenue

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/438,431A

FILING DATE: May 10, 1995

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/347,563

FILING DATE: No. 6429290ember 30, 1994

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/292,345

FILING DATE: August 17, 1994

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 600-1-087 CIP1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201 487-5800

TELEFAX: 201 343-1684

TELEX: 133521

INFORMATION FOR SEQ ID NO: 98:

SEQUENCE CHARACTERISTICS:

LENGTH: 21 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

FRAGMENT TYPE: N-terminal

US-08-438-431A-98

Query Match 3.5%; Score 15; DB 4; Length 21;  
Best Local Similarity 100.0%; Pred. No. 9.2e-08;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HHHHHSSGLVPRGS 16  
|||||  
DB 5 HHHHHSSGLVPRGS 19

RESULT 36  
US-09-270-956-46  
; Sequence 46, Application US/09270956  
; Patent No. 6451571  
; GENERAL INFORMATION:  
; APPLICANT: Loeb, Lawrence A.  
; APPLICANT: Black, Margaret E.  
; TITLE OF INVENTION: THYMIDINE KINASE MUTANTS  
; NUMBER OF SEQUENCES: 104  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: US  
; ZIP: 98104-7092

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/270,956  
FILING DATE: 17-MAR-1999  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mcmasters, David D.  
REGISTRATION NUMBER: 33,963  
REFERENCE/DOCKET NUMBER: 240052.409C3  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
TELEX: 3723836  
INFORMATION FOR SEQ ID NO: 46:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-270-956-46

Query Match 3.5%; Score 15; DB 4; Length 21;  
Best Local Similarity 100.0%; Pred. No. 9.2e-08;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HHHHHSSGLVPRGS 16  
|||||  
DB 5 HHHHHSSGLVPRGS 19

RESULT 37  
US-08-761-483-7  
; Sequence 7, Application US/08761483  
; Patent No. 6204261  
; GENERAL INFORMATION:  
; APPLICANT: Batchelor, Mark J  
; APPLICANT: Bebbington, David  
; APPLICANT: Bemis, Guy W  
; APPLICANT: Fridman, Wolf H  
; APPLICANT: Gillespie, Roger J  
; APPLICANT: Golec, Julian MC  
; APPLICANT: Gu, Yong  
; APPLICANT: Laufer, David J

APPLICANT: Livingston, David J  
APPLICANT: Matharu, Saroop S  
TITLE OF INVENTION: INHIBITORS OF INTERLEUKIN-1 BETA  
TITLE OF INVENTION: CONVERTING ENZYME  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Neave  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10020

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/761,483  
FILING DATE: 06-DEC-1996  
CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:  
NAME: Haley Jr, James F  
REGISTRATION NUMBER: 27,794  
REFERENCE/DOCKET NUMBER: VPI/96-01CIP2  
TELEPHONE: 212-596-9000  
TELEFAX: 212-596-9090  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 23 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-761-483-7

Query Match 3.5%; Score 15; DB 4; Length 23;  
Best Local Similarity 100.0%; Pred. No. 9.9e-08;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HHHHHSSGLVPRGS 16  
|||||  
DB 5 HHHHHSSGLVPRGS 19

RESULT 38  
US-08-347-563A-12  
; Sequence 12, Application US/08347563A  
; Patent No. 5935810  
; GENERAL INFORMATION:  
; APPLICANT: THE ROCKEFELLER UNIVERSITY

TITLE OF INVENTION: MODULATORS OF BODY WEIGHT, CORRESPONDING NUCLEIC  
ACIDS AND PROTEINS, AND DIAGNOSTIC AND THERAPEUTIC USES TH.  
NUMBER OF SEQUENCES: 38  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/347,563A  
FILING DATE: No. 5935810ember 30, 1994  
CLASSIFICATION: 514

;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/292,345  
;; FILING DATE: August 17, 1994  
;; CLASSIFICATION: 514  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Jackson Esq., David A.  
;; REGISTRATION NUMBER: 26,742  
;; REFERENCE/DOCKET NUMBER: 600-1-087 CIP  
;; TELEPHONE: 201 487-5800  
;; TELEFAX: 201 343-1684  
;;  
;; INFORMATION FOR SEQ ID NO: 12:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 43 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;;  
US-08-347-563A-12

Query Match 3.5%; Score 15; DB 2; Length 43;  
Best Local Similarity 100.0%; Pred. No. 1.7e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HHHHHSSGLVPRGS 16  
Db 5 HHHHHSSGLVPRGS 19

## RESULT 39

US-08-485-942A-12  
; Sequence 12, Application US/08485942A  
; Patent No. 6048837  
; GENERAL INFORMATION:  
; APPLICANT: JEFFREY M. FRIEDMAN, YIYING ZHANG, RICARDO PROENCA,  
; APPLICANT: MARGHERITA MAFFEI, JEFFREY HALAAS, KETAN GAJIWALA, AND STEPHEN K. BURLE  
; TITLE OF INVENTION: OB POLYPEPTIDE AS MODULATORS OF BODY WEIGHT (AS  
; TITLE OF INVENTION: AMENDED)  
; NUMBER OF SEQUENCES: 99  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Klauber & Jackson  
; STREET: 411 Hackensack Avenue  
; CITY: Hackensack  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/485,942A  
; FILING DATE: JUNE 7, 1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/438,431  
; FILING DATE: May 10, 1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/347,563  
; FILING DATE: August 17, 1994  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/292,345  
; FILING DATE: August 17, 1994  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jackson Esq., David A.  
; REGISTRATION NUMBER: 26,742  
; REFERENCE/DOCKET NUMBER: 600-1-087 CIP 2F  
; TELEPHONE: 201 487-5800

;; TELEFAX: 201 343-1684  
;; TELEX: 133521  
;; INFORMATION FOR SEQ ID NO: 12:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 43 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;;  
US-08-485-942A-12

Query Match 3.5%; Score 15; DB 3; Length 43;  
Best Local Similarity 100.0%; Pred. No. 1.7e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HHHHHSSGLVPRGS 16  
Db 5 HHHHHSSGLVPRGS 19

## RESULT 40

US-08-488-214A-12  
; Sequence 12, Application US/08488214A  
; Patent No. 6124439  
; GENERAL INFORMATION:  
; APPLICANT: JEFFREY M. FRIEDMAN, YIYING ZHANG, RICARDO PROENCA,  
; APPLICANT: MARGHERITA MAFFEI, JEFFREY HALAAS, KETAN GAJIWALA, AND STEPHEN K. BU  
; TITLE OF INVENTION: OB POLYPEPTIDE ANTIBODIES AND METHOD OF MAKING  
; TITLE OF INVENTION: (AS AMENDED)  
; NUMBER OF SEQUENCES: 99  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Klauber & Jackson  
; STREET: 411 Hackensack Avenue  
; CITY: Hackensack  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/488,214A  
; FILING DATE: JUNE 7, 1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/438,431  
; FILING DATE: May 10, 1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/347,563  
; FILING DATE: No. 6124439ember 30, 1994  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/292,345  
; FILING DATE: August 17, 1994  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jackson Esq., David A.  
; REGISTRATION NUMBER: 26,742  
; REFERENCE/DOCKET NUMBER: 600-1-087 CIP 2D  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201 487-5800  
; TELEFAX: 201 343-1684  
; TELEX: 133521  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 43 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
;  
US-08-488-214A-12



Query Match 3.5%; Score 15; DB 3; Length 43;  
Best Local Similarity 100.0%; Pred. No. 1.7e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HHHHHSSGLVPRGS 16  
| | | | | | | | | | | | | | | | | |  
Db 5 HHHHHSSGLVPRGS 19

Search completed: May 12, 2003, 10:28:49  
Job time : 33 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: May 12, 2003, 10:28:19 ; Search time 22 Seconds  
(without alignments)  
1802.867 Million cell updates/sec

Title: US-10-057-531A-2

Perfect score: 431

Sequence: 1 MHHHHHSSGLVPRGSMKE.....TCECTKPDYPLFDGIFCSS 431

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 349150 seqs, 92025710 residues

Word size : 0

Total number of hits satisfying chosen parameters: 349150

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Published\_Applications\_AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/CT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/CTUS\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	374	86.8	394	9 US-09-978-756-3	Sequence 3, Appli
2	231	53.6	402	9 US-10-098-514-14	Sequence 14, Appl
3	227	52.7	383	9 US-10-098-514-4	Sequence 4, Appli
4	225	52.2	383	9 US-10-098-514-2	Sequence 2, Appli
5	60	13.9	95	10 US-09-134-333-2	Sequence 2, Appli
6	60	13.9	108	10 US-09-134-333-10	Sequence 10, Appl
7	60	13.9	116	10 US-09-134-333-5	Sequence 5, Appli
8	60	13.9	127	10 US-09-134-333-8	Sequence 8, Appli
9	50	11.6	516	10 US-09-804-626-4	Sequence 4, Appli
10	50	11.6	518	10 US-09-804-626-2	Sequence 2, Appli
11	50	11.6	824	9 US-10-046-583A-3	Sequence 3, Appli
12	46	10.7	96	9 US-09-978-756-1	Sequence 1, Appli
13	46	10.7	114	9 US-10-087-464-35	Sequence 35, Appl
14	46	10.7	201	9 US-09-681-938-1	Sequence 1, Appli
15	46	10.7	376	9 US-09-978-756-2	Sequence 2, Appli
16	46	10.7	376	9 US-10-087-464-34	Sequence 34, Appl
17	46	10.7	378	9 US-10-087-464-11	Sequence 11, Appl
18	46	10.7	1639	9 US-10-087-464-10	Sequence 10, Appl
19	45	10.4	355	9 US-10-012-896-1011	Sequence 1011, Ap

32	7.4	55	9	US-10-042-945-28	Sequence 28, Appl
32	7.4	55	10	US-09-757-417-28	Sequence 28, Appl
21	3.5	15	9	US-09-872-712-1	Sequence 1, Appli
23	3.5	18	10	US-09-905-831-1	Sequence 5, Appli
24	3.5	20	9	US-09-320-907B-5	Sequence 114, App
25	3.5	20	10	US-09-903-412-114	Sequence 5, Appli
26	3.5	21	9	US-09-096-749A-114	Sequence 98, Appl
27	3.5	21	10	US-10-060-275-5	Sequence 12, Appl
28	3.5	43	10	US-09-736-084-98	Sequence 12, Appl
29	3.5	43	10	US-09-736-084-12	Sequence 8, Appli
30	3.5	121	9	US-10-267-311-12	Sequence 19, Appl
31	3.5	124	9	US-09-981-286A-8	Sequence 23, Appl
32	3.5	149	9	US-09-876-348A-19	Sequence 27, Appl
33	3.5	149	9	US-09-876-348A-23	Sequence 31, Appl
34	3.5	149	9	US-09-876-348A-27	Sequence 35, Appl
35	3.5	149	9	US-09-876-348A-31	Sequence 39, Appl
36	3.5	149	9	US-09-876-348A-35	Sequence 19, Appl
37	3.5	149	9	US-09-876-348A-39	Sequence 23, Appl
38	3.5	149	9	US-09-876-796A-19	Sequence 27, Appl
39	3.5	149	9	US-09-876-796A-23	Sequence 31, Appl
40	3.5	149	9	US-09-876-796A-27	Sequence 35, Appl
41	3.5	149	9	US-09-876-796A-31	Sequence 39, Appl
42	3.5	149	9	US-09-876-796A-35	Sequence 2, Appli
43	3.5	149	9	US-09-876-796A-39	Sequence 4, Appli
44	3.5	153	10	US-09-798-338-2	Sequence 6, Appli
45	3.5	157	10	US-09-798-338-4	Sequence 8, Appli
46	3.5	163	10	US-09-798-338-6	Sequence 4, Appli
47	3.5	167	10	US-09-798-338-8	Sequence 4, Appli
48	3.5	170	9	US-10-068-059-4	Sequence 17, Appl
49	3.5	173	9	US-09-876-348A-17	Sequence 21, Appl
50	3.5	173	9	US-09-876-348A-21	Sequence 29, Appl
51	3.5	173	9	US-09-876-348A-29	Sequence 33, Appl
52	3.5	173	9	US-09-876-348A-33	Sequence 37, Appl
53	3.5	173	9	US-09-876-348A-37	Sequence 17, Appl
54	3.5	173	9	US-09-876-796A-17	Sequence 21, Appl
55	3.5	173	9	US-09-876-796A-21	Sequence 29, Appl
56	3.5	173	9	US-09-876-796A-33	Sequence 37, Appl
57	3.5	173	9	US-09-876-796A-37	Sequence 25, Appl
58	3.5	173	9	US-09-876-796A-39	Sequence 25, Appl
59	3.5	174	9	US-09-876-348A-25	Sequence 2, Appli
60	3.5	174	9	US-09-876-796A-25	Sequence 6, Appli
61	3.5	299	10	US-09-850-677A-2	Sequence 16, Appl
62	3.5	333	9	US-10-108-195-6	Sequence 1, Appli
63	3.5	371	9	US-09-849-243-16	Sequence 2239, Ap
64	3.5	386	9	US-09-948-193-1	Sequence 121, App
65	3.5	391	9	US-10-046-935-2239	Sequence 55, Appl
66	3.5	391	9	US-10-146-502-2239	Sequence 6, Appli
67	3.5	398	1	US-08-979-847-121	Sequence 21, Appl
68	3.5	409	9	US-10-267-311-55	Sequence 12, Appl
69	3.5	709	9	US-10-068-059-8	Sequence 13, Appl
70	3.5	746	9	US-10-068-059-6	Sequence 9, Appli
71	3.5	948	9	US-10-267-311-21	Sequence 8, Appli
72	3.2	948	9	US-09-849-243-17	Sequence 5, Appli
73	3.2	19	9	US-09-849-243-12	Sequence 12, Appl
74	2.1	13	9	US-10-108-195-13	Sequence 16, Appl
75	2.1	22	9	US-09-331-631A-16	Sequence 9, Appli
76	2.1	24	10	US-10-112-193-9	Sequence 8, Appli
77	2.1	24	10	US-09-934-465-8	Sequence 9, Appli
78	2.1	24	10	US-09-884-733-9	Sequence 5, Appli
79	2.1	24	12	US-10-080-455-5	Sequence 6, Appli
80	2.1	131	12	US-10-016-447-6	Sequence 40, Appl
81	2.1	135	10	US-09-804-615-40	Sequence 2, Appli
82	2.1	183	9	US-09-832-659-4	Sequence 6, Appli
83	2.1	183	9	US-09-832-658-2	Sequence 11, Appl
84	2.1	195	10	US-09-949-780-6	Sequence 32, Appl
85	2.1	600	9	US-09-930-125-11	Sequence 13, Appl
86	1.9	14	9	US-10-004-381-32	Sequence 4, Appli
87	1.9	31	9	US-09-996-069-12	Sequence 13, Appl
88	1.9	60	9	US-09-820-843A-13	Sequence 4, Appli
89	1.9	244	9	US-09-899-954B-4	Sequence 6, Appli
90	1.9	244	9	US-09-899-954B-6	Sequence 332, App
91	1.9	342	9	US-10-125-635A-332	Sequence 332, App
92	1.9	342	9	US-09-938-864-332	





385	1.4	14	1	US-08-464-363-76	Sequence 76, Appl	458	6	1.4	27	9	US-10-226-956-283	Sequence 283, App
386	1.4	14	9	US-09-784-199-9	Sequence 9, Appl	459	6	1.4	27	9	US-10-226-956-289	Sequence 289, App
387	1.4	14	9	US-10-196-107A-58	Sequence 58, Appl	460	6	1.4	27	10	US-09-846-729A-22	Sequence 22, Appl
388	1.4	14	10	US-09-374-671-58	Sequence 58, Appl	461	6	1.4	27	10	US-09-846-729A-26	Sequence 26, Appl
389	1.4	14	10	US-09-823-266-1	Sequence 1, Appl	462	6	1.4	27	10	US-09-898-461-9	Sequence 9, Appl
390	1.4	14	10	US-09-900-530A-19	Sequence 19, Appl	463	6	1.4	28	9	US-10-750-152	Sequence 152, App
391	1.4	15	9	US-10-014-485A-43	Sequence 43, Appl	464	6	1.4	29	10	US-10-282-121-12	Sequence 12, Appl
392	1.4	15	9	US-10-010-184A-4	Sequence 4, Appl	465	6	1.4	29	10	US-09-880-901-2	Sequence 2, Appl
393	1.4	15	9	US-10-196-107A-53	Sequence 53, Appl	466	6	1.4	29	10	US-09-935-291A-24	Sequence 24, Appl
394	1.4	15	9	US-09-872-712-2	Sequence 2, Appl	467	6	1.4	30	9	US-09-911-569-98	Sequence 98, Appl
395	1.4	15	9	US-09-872-712-28	Sequence 28, Appl	468	6	1.4	30	9	US-09-911-569-100	Sequence 100, App
396	1.4	15	9	US-10-174-105A-43	Sequence 43, Appl	469	6	1.4	30	9	US-09-911-569-104	Sequence 104, App
397	1.4	15	10	US-09-374-671-53	Sequence 53, Appl	470	6	1.4	30	10	US-09-864-761-36251	Sequence 36251, A
398	1.4	15	10	US-09-821-883-16	Sequence 16, Appl	471	6	1.4	30	10	US-09-754-826-3	Sequence 3, Appl
399	1.4	15	12	US-10-139-841-43	Sequence 43, Appl	472	6	1.4	31	9	US-09-259-658-58	Sequence 58, Appl
400	1.4	16	9	US-09-911-569-103	Sequence 103, App	473	6	1.4	31	9	US-09-996-069-4	Sequence 4, Appl
401	1.4	16	9	US-09-911-569-105	Sequence 105, App	474	6	1.4	31	10	US-09-925-442-37	Sequence 37, Appl
402	1.4	16	10	US-09-809-517A-4	Sequence 4, Appl	475	6	1.4	32	9	US-10-174-410-182	Sequence 182, App
403	1.4	16	10	US-09-809-517A-29	Sequence 29, Appl	476	6	1.4	32	10	US-09-864-761-34706	Sequence 34706, A
404	1.4	16	10	US-09-351-794A-7	Sequence 7, Appl	477	6	1.4	33	9	US-09-259-658-49	Sequence 49, Appl
405	1.4	16	10	US-09-948-018-25	Sequence 25, Appl	478	6	1.4	33	9	US-09-259-658-50	Sequence 50, Appl
406	1.4	17	9	US-09-051-013-6	Sequence 6, Appl	479	6	1.4	33	9	US-09-866-538-21	Sequence 21, Appl
407	1.4	17	9	US-10-083-590-12	Sequence 12, Appl	480	6	1.4	33	10	US-10-121-258-22	Sequence 22, Appl
408	1.4	17	9	US-09-925-287-1	Sequence 1, Appl	481	6	1.4	33	10	US-09-864-761-36853	Sequence 36853, A
409	1.4	17	9	US-10-047-991-8	Sequence 8, Appl	482	6	1.4	35	9	US-10-083-815-68	Sequence 68, Appl
410	1.4	17	10	US-09-809-517A-32	Sequence 32, Appl	483	6	1.4	35	10	US-09-864-761-36426	Sequence 36426, A
411	1.4	18	1	US-08-954-771-47	Sequence 47, Appl	484	6	1.4	36	9	US-09-876-904A-309	Sequence 309, App
412	1.4	18	9	US-10-105-717-3	Sequence 3, Appl	485	6	1.4	36	9	US-09-276-455-8	Sequence 8, Appl
413	1.4	18	9	US-09-876-904A-92	Sequence 92, Appl	486	6	1.4	37	9	US-09-575-847-5	Sequence 5, Appl
414	1.4	18	10	US-09-809-517A-1	Sequence 1, Appl	487	6	1.4	37	10	US-09-911-569-97	Sequence 97, Appl
415	1.4	18	10	US-09-864-761-34373	Sequence 34373, A	488	6	1.4	37	10	US-09-864-761-40909	Sequence 40909, A
416	1.4	18	10	US-09-864-761-37659	Sequence 37659, A	489	6	1.4	39	9	US-09-996-069-1	Sequence 1, Appl
417	1.4	19	9	US-09-933-999A-31	Sequence 31, Appl	490	6	1.4	39	9	US-10-092-154-928	Sequence 928, App
418	1.4	19	9	US-09-996-069-13	Sequence 13, Appl	491	6	1.4	39	9	US-10-100-679-78	Sequence 78, Appl
419	1.4	19	9	US-09-876-904A-627	Sequence 627, App	492	6	1.4	39	10	US-10-102-806-429	Sequence 429, App
420	1.4	20	9	US-09-051-013-3	Sequence 3, Appl	493	6	1.4	39	10	US-09-884-681-39	Sequence 39, Appl
421	1.4	20	9	US-08-300-425B-26	Sequence 26, Appl	494	6	1.4	39	10	US-09-764-847-928	Sequence 928, App
422	1.4	21	8	US-08-326-626-13	Sequence 13, Appl	495	6	1.4	42	9	US-09-972-708-2	Sequence 2, Appl
423	1.4	21	9	US-10-082-659-16	Sequence 16, Appl	496	6	1.4	42	10	US-09-903-452-12	Sequence 12, Appl
424	1.4	21	9	US-09-092-296-20	Sequence 20, Appl	497	6	1.4	43	9	US-09-259-658-1	Sequence 1, Appl
425	1.4	21	9	US-09-104-408-37	Sequence 37, Appl	498	6	1.4	43	9	US-09-259-658-25	Sequence 25, Appl
426	1.4	21	9	US-10-216-408-27	Sequence 27, Appl	499	6	1.4	43	9	US-09-259-658-27	Sequence 27, Appl
427	1.4	21	9	US-09-939-126-8	Sequence 8, Appl	500	6	1.4	43	9	US-09-259-658-30	Sequence 30, Appl
428	1.4	21	9	US-09-939-126-9	Sequence 9, Appl	501	6	1.4	43	9	US-09-259-658-38	Sequence 38, Appl
429	1.4	21	9	US-09-939-126-10	Sequence 10, Appl	502	6	1.4	43	9	US-09-259-658-40	Sequence 40, Appl
430	1.4	21	9	US-09-939-126-11	Sequence 11, Appl	503	6	1.4	43	9	US-09-259-658-42	Sequence 42, Appl
431	1.4	21	9	US-09-939-126-12	Sequence 12, Appl	504	6	1.4	43	9	US-09-259-658-45	Sequence 45, Appl
432	1.4	21	9	US-10-278-547-49	Sequence 49, Appl	505	6	1.4	43	9	US-09-259-658-48	Sequence 48, Appl
433	1.4	21	10	US-09-050-516-49	Sequence 49, Appl	506	6	1.4	43	9	US-09-259-658-52	Sequence 52, Appl
434	1.4	21	10	US-09-276-600-11	Sequence 11, Appl	507	6	1.4	43	9	US-09-259-658-54	Sequence 54, Appl
435	1.4	21	10	US-09-099-823-27	Sequence 27, Appl	508	6	1.4	43	9	US-09-259-658-56	Sequence 56, Appl
436	1.4	21	10	US-09-234-717-25	Sequence 25, Appl	509	6	1.4	43	10	US-09-864-761-46454	Sequence 46454, A
437	1.4	21	10	US-09-850-178-19	Sequence 19, Appl	510	6	1.4	44	10	US-09-911-569-99	Sequence 99, Appl
438	1.4	21	10	US-09-193-538-23	Sequence 23, Appl	511	6	1.4	44	10	US-09-864-761-39281	Sequence 39281, A
439	1.4	21	10	US-09-250-883-23	Sequence 23, Appl	512	6	1.4	44	10	US-09-766-378A-34	Sequence 34, Appl
440	1.4	21	10	US-09-096-259-33	Sequence 33, Appl	513	6	1.4	45	9	US-09-921-144-6	Sequence 6, Appl
441	1.4	21	10	US-09-215-652-48	Sequence 48, Appl	514	6	1.4	45	10	US-09-766-378A-33	Sequence 33, Appl
442	1.4	21	10	US-09-347-064-18	Sequence 18, Appl	515	6	1.4	46	9	US-09-996-069-5	Sequence 5, Appl
443	1.4	21	10	US-09-049-695A-22	Sequence 22, Appl	516	6	1.4	46	9	US-09-764-891-4755	Sequence 4755, Ap
444	1.4	21	10	US-09-092-297-22	Sequence 22, Appl	517	6	1.4	47	9	US-09-801-089-18	Sequence 18, Appl
445	1.4	21	12	US-10-025-167-51	Sequence 51, Appl	518	6	1.4	47	9	US-10-173-570-5	Sequence 5, Appl
446	1.4	22	9	US-09-911-569-102	Sequence 102, App	519	6	1.4	47	10	US-09-738-945-5	Sequence 5, Appl
447	1.4	23	9	US-10-105-717-5	Sequence 5, Appl	520	6	1.4	47	10	US-09-923-243-1	Sequence 1, Appl
448	1.4	23	9	US-10-171-417-5	Sequence 5, Appl	521	6	1.4	48	9	US-09-921-144-8	Sequence 8, Appl
449	1.4	23	9	US-10-171-417-6	Sequence 6, Appl	522	6	1.4	48	10	US-09-864-761-37882	Sequence 37882, A
450	1.4	23	9	US-09-939-126-7	Sequence 7, Appl	523	6	1.4	51	9	US-09-996-069-2	Sequence 2, Appl
451	1.4	23	9	US-09-876-904A-624	Sequence 624, App	524	6	1.4	51	10	US-09-896-720-22	Sequence 22, Appl
452	1.4	24	9	US-09-876-904A-623	Sequence 623, App	525	6	1.4	51	10	US-09-864-761-38187	Sequence 38187, A
453	1.4	24	10	US-09-841-133-221	Sequence 221, App	526	6	1.4	52	10	US-09-864-761-42814	Sequence 42814, A
454	1.4	25	9	US-09-911-569-96	Sequence 96, Appl	527	6	1.4	54	9	US-09-986-480-209	Sequence 209, App
455	1.4	25	10	US-09-864-761-45218	Sequence 45218, A	528	6	1.4	54	10	US-09-996-069-3	Sequence 3, Appl
456	1.4	26	10	US-09-773-385-18	Sequence 18, Appl	529	6	1.4	54	10	US-09-925-302-817	Sequence 817, App
457	1.4	26	10	US-09-864-761-47310	Sequence 47310, A	530	6	1.4	54	10	US-09-864-761-38540	Sequence 38540, A









969 6 1.4 357 9 US-09-895-814-826 Sequence 826, App  
970 6 1.4 357 10 US-09-759-143-826 Sequence 826, App  
971 6 1.4 357 10 US-09-780-669-826 Sequence 826, App  
972 6 1.4 357 10 US-09-822-827-826 Sequence 826, App  
973 6 1.4 358 10 US-09-801-368-148 Sequence 148, App  
974 6 1.4 359 10 US-09-764-864-996 Sequence 996, App  
975 6 1.4 361 9 US-09-839-894-10 Sequence 10, Appl  
976 6 1.4 361 9 US-09-839-894-28 Sequence 28, Appl  
977 6 1.4 363 9 US-10-159-151-10 Sequence 10, Appl  
978 6 1.4 364 9 US-09-829-275-1 Sequence 1, Appl  
979 6 1.4 366 10 US-09-887-569A-2 Sequence 2, Appl  
980 6 1.4 368 9 US-10-084-700-30 Sequence 30, Appl  
981 6 1.4 370 9 US-09-828-323A-74 Sequence 74, Appl  
982 6 1.4 371 9 US-10-000-903-6 Sequence 6, Appl  
983 6 1.4 371 10 US-09-850-716A-433 Sequence 433, App  
984 6 1.4 371 10 US-09-897-778-433 Sequence 433, App  
985 6 1.4 372 9 US-09-813-718-4 Sequence 4, Appl  
986 6 1.4 372 9 US-09-813-718-8 Sequence 8, Appl  
987 6 1.4 373 9 US-10-159-151-4 Sequence 10, Appl  
988 6 1.4 375 10 US-09-764-864-1316 Sequence 1316, App  
989 6 1.4 378 1 US-08-979-847-122 Sequence 122, App  
990 6 1.4 378 10 US-09-764-864-863 Sequence 863, App  
991 6 1.4 379 10 US-09-134-333-11 Sequence 11, Appl  
992 6 1.4 381 9 US-10-012-896-941 Sequence 941, App  
993 6 1.4 381 9 US-09-895-793-941 Sequence 941, App  
994 6 1.4 381 9 US-09-895-814-941 Sequence 941, App  
995 6 1.4 381 9 US-10-159-151-6 Sequence 6, Appl  
996 6 1.4 381 10 US-09-780-669-941 Sequence 941, App  
997 6 1.4 381 10 US-09-822-827-941 Sequence 941, App  
998 6 1.4 383 9 US-10-000-903-23 Sequence 23, Appl  
999 6 1.4 383 10 US-09-883-096-5 Sequence 5, Appl  
1000 6 1.4 387 9 US-09-972-268-16 Sequence 16, Appl

ALIGNMENTS

RESULT 1  
US-09-978-756-3  
; Sequence 3, Application US/09978756  
; Patent No. US20020160017A1  
; GENERAL INFORMATION:  
; APPLICANT: Holder, Anthony  
; APPLICANT: Birdsell, Berry  
; APPLICANT: Feeney, James  
; APPLICANT: Morgan, William  
; APPLICANT: Syed, Shabih  
; TITLE OF INVENTION: Malaria Vaccine  
; FILE REFERENCE: 18396/1005  
; CURRENT APPLICATION NUMBER: US/09/978,756  
; CURRENT FILING DATE: 2001-10-16  
; PRIOR APPLICATION NUMBER: PCT/GB00/01558  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR FILING DATE: 1999-05-13  
; PRIOR FILING DATE: 1999-05-13  
; PRIOR FILING DATE: 1999-05-25  
; PRIOR APPLICATION NUMBER: 2,271,451  
; PRIOR FILING DATE: 1999-05-25  
; PRIOR APPLICATION NUMBER: 9909072.2  
; PRIOR FILING DATE: 1999-04-20  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 3  
; LENGTH: 394  
; TYPE: PRT  
; ORGANISM: Plasmodium falciparum  
US-09-978-756-3

Query Match 86.8%; Score 374; DB 9; Length 394;  
Best Local Similarity 100.0%; Pred. No. 3.9e-293;  
Matches 374; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 58 AISVTMDNILSGFENEYDVILKPLAGVYRSILKKQIEKNFTFNLNLDILNSRLKRRKY 117  
|||||

Db 1 AISVTMDNILSGFENEYDVILKPLAGVYRSILKKQIEKNFTFNLNLDILNSRLKRRKY 60  
QY 118 FLDVLESLDMQFKHISSENEYIIEDSFKLLNSEQNTLLKSKYIKESVENDIKKFAQEGIS 177  
|||||  
Db 61 FLDVLESLDMQFKHISSENEYIIEDSFKLLNSEQNTLLKSKYIKESVENDIKKFAQEGIS 120  
QY 178 YYEKVLAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAKTDEQKESKFLPFLTNIETL 237  
|||||  
Db 121 YYEKVLAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAKTDEQKESKFLPFLTNIETL 180  
QY 238 YNNLVNKIDDDYILNLIKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNPYDEFAIKK 297  
|||||  
Db 181 YNNLVNKIDDDYILNLIKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNPYDEFAIKK 240  
QY 298 LINDDTKMDLGLKLLSTGLVQNFNTIISKLEIEGFQDMLNISQHCVKKQCPNSGCFR 357  
|||||  
Db 241 LINDDTKMDLGLKLLSTGLVQNFNTIISKLEIEGFQDMLNISQHCVKKQCPNSGCFR 300  
QY 358 HLDRECECKLLNYKQEGDKCVENPPTCENNNGCGDADATCTEEDSGSSRKKITCECTK 417  
|||||  
Db 301 HLDRECECKLLNYKQEGDKCVENPPTCENNNGCGDADATCTEEDSGSSRKKITCECTK 360  
QY 418 PDSYPLFDGIEFCSS 431  
Db 361 PDSYPLFDGIEFCSS 374  
RESULT 2  
US-10-098-514-14  
; Sequence 14, Application US/10098514  
; Publication No. US20020194648A1  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Sandra P  
; APPLICANT: Christopher, David A  
; APPLICANT: Vine, Benjamin  
; APPLICANT: Su, Wei-Wen  
; APPLICANT: Bugos, Robert  
; TITLE OF INVENTION: PLASMODIUM FALCIPARUM MEROZOITE SURFACE PROTEIN-1 MALARIA PROD  
; FILE REFERENCE: A-71339/RFT/TAL/NBC  
; CURRENT APPLICATION NUMBER: US/10/098,514  
; CURRENT FILING DATE: 2002-08-06  
; PRIOR APPLICATION NUMBER: US 09/500,376  
; PRIOR FILING DATE: 2000-02-08  
; PRIOR APPLICATION NUMBER: US 60/274,599  
; PRIOR FILING DATE: 2001-03-09  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 14  
; LENGTH: 402  
; TYPE: PRT  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: synthetic  
US-10-098-514-14

Query Match 53.6%; Score 231; DB 9; Length 402;  
Best Local Similarity 100.0%; Pred. No. 4.3e-178;  
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 58 AISVTMDNILSGFENEYDVILKPLAGVYRSILKKQIEKNFTFNLNLDILNSRLKRRKY 117  
|||||  
Db 22 AISVTMDNILSGFENEYDVILKPLAGVYRSILKKQIEKNFTFNLNLDILNSRLKRRKY 81  
QY 118 FLDVLESLDMQFKHISSENEYIIEDSFKLLNSEQNTLLKSKYIKESVENDIKKFAQEGIS 177  
|||||  
Db 82 FLDVLESLDMQFKHISSENEYIIEDSFKLLNSEQNTLLKSKYIKESVENDIKKFAQEGIS 141  
QY 178 YYEKVLAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAKTDEQKESKFLPFLTNIETL 237  
|||||  
Db 142 YYEKVLAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAKTDEQKESKFLPFLTNIETL 201  
QY 238 YNNLVNKIDDDYILNLIKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKN 288

```
Db 202 YNNLVNKIDDYLNKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFKN 252
|||||
RESULT 3
US-10-098-514-4
; Sequence 4, Application US/10098514
; Publication No. US20020194648A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Sandra P
; APPLICANT: Christopher, David A
; APPLICANT: Vine, Benjamin
; APPLICANT: Su, Wei-Wen
; APPLICANT: Bugos, Robert
; TITLE OF INVENTION: PLASMODIUM FALCIPARUM MEROZOITE SURFACE PROTEIN-1 MALARIA PRODU
; FILE REFERENCE: A-71339/REF/TAL/NEC
; CURRENT APPLICATION NUMBER: US/10/098,514
; CURRENT FILING DATE: 2002-08-06
; PRIOR APPLICATION NUMBER: US 09/500,376
; PRIOR FILING DATE: 2000-02-08
; PRIOR APPLICATION NUMBER: US 60/274,599
; PRIOR FILING DATE: 2001-03-09
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: synthetic
; NAME/KEY: MISC_FEATURE
; LOCATION: (380)..(380)
; OTHER INFORMATION: "xaa" at position 380 represents a stop codon
US-10-098-514-4
Query Match 52.7%; Score 227; DB 9; Length 383;
Best Local Similarity 100.0%; Pred. No. 6.9e-175;
Matches 227; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 62 TMDNILSGFENEYDVYILKPLAGVYRSLLKQIEKNITFTNLDILNSRLKRRKYFLDV 121
Db 7 TMDNILSGFENEYDVYILKPLAGVYRSLLKQIEKNITFTNLDILNSRLKRRKYFLDV 66
QY 122 LESDLMQFKHISSEYIIEDSFKLLNSEQNTLLKSKYIKESVENDIKFAQEGISYIEK 181
Db 67 LESDLMQFKHISSEYIIEDSFKLLNSEQNTLLKSKYIKESVENDIKFAQEGISYIEK 126
QY 182 VLAKYKDDLESIKKVIKEEKEFPSPPTPPSPAKTDEQKESKFLPFLTNIETLYNNL 241
Db 127 VLAKYKDDLESIKKVIKEEKEFPSPPTPPSPAKTDEQKESKFLPFLTNIETLYNNL 186
QY 242 VNKIDDYLNKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFKN 288
Db 187 VNKIDDYLNKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFKN 233
RESULT 4
US-10-098-514-2
; Sequence 2, Application US/10098514
; Publication No. US20020194648A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Sandra P
; APPLICANT: Christopher, David A
; APPLICANT: Vine, Benjamin
; APPLICANT: Su, Wei-Wen
; APPLICANT: Bugos, Robert
; TITLE OF INVENTION: PLASMODIUM FALCIPARUM MEROZOITE SURFACE PROTEIN-1 MALARIA PRODU
; FILE REFERENCE: A-71339/REF/TAL/NEC
; CURRENT APPLICATION NUMBER: US/10/098,514
; CURRENT FILING DATE: 2002-08-06
```

```
; PRIOR APPLICATION NUMBER: US 09/500,376
; PRIOR FILING DATE: 2000-02-08
; PRIOR APPLICATION NUMBER: US 60/274,599
; PRIOR FILING DATE: 2001-03-09
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: synthetic
; NAME/KEY: MISC_FEATURE
; LOCATION: (380)..(380)
; OTHER INFORMATION: "xaa" at position 380 represents a stop codon
US-10-098-514-2
Query Match 52.2%; Score 225; DB 9; Length 383;
Best Local Similarity 100.0%; Pred. No. 2.8e-173;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 64 DNILSGFENEYDVYILKPLAGVYRSLLKQIEKNITFTNLDILNSRLKRRKYFLDVLE 123
Db 9 DNILSGFENEYDVYILKPLAGVYRSLLKQIEKNITFTNLDILNSRLKRRKYFLDVLE 68
QY 124 SLDLMQFKHISSEYIIEDSFKLLNSEQNTLLKSKYIKESVENDIKFAQEGISYIEKVL 183
Db 69 SLDLMQFKHISSEYIIEDSFKLLNSEQNTLLKSKYIKESVENDIKFAQEGISYIEKVL 128
QY 184 AKYKDDLESIKKVIKEEKEFPSPPTPPSPAKTDSQKESKFLPFLTNIETLYNNLVN 243
Db 129 AKYKDDLESIKKVIKEEKEFPSPPTPPSPAKTDSQKESKFLPFLTNIETLYNNLVN 188
QY 244 KIDDYLNKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFKN 288
Db 189 KIDDYLNKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFKN 233
RESULT 5
US-09-134-333-2
; Sequence 2, Application US/09134333
; Patent No. US20020076403A1
; GENERAL INFORMATION:
; APPLICANT: LONGACRE-ANDRE, SHIRLEY
; APPLICANT: ROTH, CHARLES
; APPLICANT: NATO, FARIDABANO
; APPLICANT: BARNWELL, JOHN
; APPLICANT: MENDIS, KAMINI
; TITLE OF INVENTION: RECOMBINANT PROTEIN CONTAINING A C-TERMINAL FRAGMENT OF
; FILE REFERENCE: 0660-0135-0XCIP
; CURRENT APPLICATION NUMBER: US/09/134,333
; CURRENT FILING DATE: 1999-04-18
; EARLIER APPLICATION NUMBER: PCT/FR97/00290
; EARLIER FILING DATE: 1997-02-14
; EARLIER APPLICATION NUMBER: FR96/01822
; EARLIER FILING DATE: 1996-02-14
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 95
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-134-333-2
Query Match 13.9%; Score 60; DB 10; Length 95;
Best Local Similarity 100.0%; Pred. No. 4.8e-41;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 338 NISQHCQVKKQCPENSGCFRHLDEREECKILLNYKQESDKCVENPNPTCNENGGCDADA 397
```

Db 3 NISQHCYKQCPENSGCFRHLDERECKCLLNKYQEGDKCVENPNTCENNNGCDADA 62  
|||||

RESULT 6  
US-09-134-333-10  
; Sequence 10, Application US/09134333  
; Patent No. US20020076403A1  
; GENERAL INFORMATION:  
; APPLICANT: LONGACRE-ANDRE, SHIRLEY  
; APPLICANT: ROTH, CHARLES  
; APPLICANT: NATO, FARIDABANO  
; APPLICANT: BARNWELL, JOHN  
; APPLICANT: MENDIS, KAMINI  
; TITLE OF INVENTION: RECOMBINANT PROTEIN CONTAINING A C-TERMINAL FRAGMENT OF  
; FILE REFERENCE: 0660-0135-0XCIP  
; CURRENT APPLICATION NUMBER: US/09/134,333  
; CURRENT FILING DATE: 1999-04-18  
; EARLIER APPLICATION NUMBER: PCT/FR97/00290  
; EARLIER FILING DATE: 1997-02-14  
; EARLIER APPLICATION NUMBER: FR96/01822  
; EARLIER FILING DATE: 1996-02-14  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 10  
; LENGTH: 108  
; TYPE: PRT  
; ORGANISM: Plasmodium falciparum  
US-09-134-333-10

Query Match 13.9%; Score 60; DB 10; Length 108;  
Best Local Similarity 100.0%; Pred. No. 5.3e-41;  
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 338 NISQHCYKQCPENSGCFRHLDERECKCLLNKYQEGDKCVENPNTCENNNGCDADA 397  
|||||

Db 16 NISQHCYKQCPENSGCFRHLDERECKCLLNKYQEGDKCVENPNTCENNNGCDADA 75  
|||||

RESULT 7  
US-09-134-333-5  
; Sequence 5, Application US/09134333  
; Patent No. US20020076403A1  
; GENERAL INFORMATION:  
; APPLICANT: LONGACRE-ANDRE, SHIRLEY  
; APPLICANT: ROTH, CHARLES  
; APPLICANT: NATO, FARIDABANO  
; APPLICANT: BARNWELL, JOHN  
; APPLICANT: MENDIS, KAMINI  
; TITLE OF INVENTION: RECOMBINANT PROTEIN CONTAINING A C-TERMINAL FRAGMENT OF  
; FILE REFERENCE: 0660-0135-0XCIP  
; CURRENT APPLICATION NUMBER: US/09/134,333  
; CURRENT FILING DATE: 1999-04-18  
; EARLIER APPLICATION NUMBER: PCT/FR97/00290  
; EARLIER FILING DATE: 1997-02-14  
; EARLIER APPLICATION NUMBER: FR96/01822  
; EARLIER FILING DATE: 1996-02-14  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 116  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC  
US-09-134-333-5

Query Match 13.9%; Score 60; DB 10; Length 116;  
Best Local Similarity 100.0%; Pred. No. 5.6e-41;  
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 338 NISQHCYKQCPENSGCFRHLDERECKCLLNKYQEGDKCVENPNTCENNNGCDADA 397  
|||||

Db 3 NISQHCYKQCPENSGCFRHLDERECKCLLNKYQEGDKCVENPNTCENNNGCDADA 62  
|||||

RESULT 8  
US-09-134-333-8  
; Sequence 8, Application US/09134333  
; Patent No. US20020076403A1  
; GENERAL INFORMATION:  
; APPLICANT: LONGACRE-ANDRE, SHIRLEY  
; APPLICANT: ROTH, CHARLES  
; APPLICANT: NATO, FARIDABANO  
; APPLICANT: BARNWELL, JOHN  
; APPLICANT: MENDIS, KAMINI  
; TITLE OF INVENTION: RECOMBINANT PROTEIN CONTAINING A C-TERMINAL FRAGMENT OF  
; FILE REFERENCE: 0660-0135-0XCIP  
; CURRENT APPLICATION NUMBER: US/09/134,333  
; CURRENT FILING DATE: 1999-04-18  
; EARLIER APPLICATION NUMBER: PCT/FR97/00290  
; EARLIER FILING DATE: 1997-02-14  
; EARLIER APPLICATION NUMBER: FR96/01822  
; EARLIER FILING DATE: 1996-02-14  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 8  
; LENGTH: 127  
; TYPE: PRT  
; ORGANISM: Plasmodium falciparum  
US-09-134-333-8

Query Match 13.9%; Score 60; DB 10; Length 127;  
Best Local Similarity 100.0%; Pred. No. 6.1e-41;  
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 338 NISQHCYKQCPENSGCFRHLDERECKCLLNKYQEGDKCVENPNTCENNNGCDADA 397  
|||||

Db 35 NISQHCYKQCPENSGCFRHLDERECKCLLNKYQEGDKCVENPNTCENNNGCDADA 94  
|||||

RESULT 9  
US-09-804-626-4  
; Sequence 4, Application US/09804626  
; Patent No. US20020128190A1  
; GENERAL INFORMATION:  
; APPLICANT: Lustbader, Joyce  
; TITLE OF INVENTION: EXPRESSION OF PROPERLY FOLDED AND SOLUBLE EXTRACELLULAR DOMAIN  
; FILE REFERENCE: 0575/62259/JPW/SHS  
; CURRENT APPLICATION NUMBER: US/09/804,626  
; CURRENT FILING DATE: 2001-03-09  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 4  
; LENGTH: 516  
; TYPE: PRT  
; ORGANISM: Homo Sapiens  
US-09-804-626-4

Query Match 11.6%; Score 50; DB 10; Length 516;  
Best Local Similarity 100.0%; Pred. No. 2.2e-32;  
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHHHHISSGLVPRGSGMKETAAAKFERQHMDSPDLGTDGDDDDKAMADIGS 50  
|||||

Db 116 MHHHHISSGLVPRGSGMKETAAAKFERQHMDSPDLGTDGDDDDKAMADIGS 165  
|||||

RESULT 10  
US-09-804-626-2  
; Sequence 2, Application US/09804626

```
; Patent No. US20020128190A1
; GENERAL INFORMATION:
; APPLICANT: Lobel, Leslie
; APPLICANT: Lustbader, Joyce
; TITLE OF INVENTION: EXPRESSION OF PROPERLY FOLDED AND SOLUBLE EXTRACELLULAR DOMAIN OF
; TITLE OF INVENTION: GONADOTROPIN RECEPTOR
; FILE REFERENCE: 0575/62259/JPW/SHS
; CURRENT APPLICATION NUMBER: US/09/804,626
; CURRENT FILING DATE: 2001-03-09
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-09-804-626-2

Query Match      11.6%; Score 50; DB 10; Length 518;
Best Local Similarity 100.0%; Pred. No. 2.2e-32;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHSSGLVPRGSGMKETAARKFERQHMDSPDLGTDDDDKAMADIGS 50
|||||
DB 116 MHHHHHSSGLVPRGSGMKETAARKFERQHMDSPDLGTDDDDKAMADIGS 165

RESULT 11
US-10-046-583A-3
; Sequence 3, Application US/10046583A
; Patent No. US20020168743A1
; GENERAL INFORMATION:
; APPLICANT: Rice, John
; APPLICANT: Kloti, Andreas
; APPLICANT: Crawford, John
; APPLICANT: Lanning, Beth
; APPLICANT: Stewart, Sandy
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
; TITLE OF INVENTION: IDENTIFICATION OF MODULATORS OF DEOXYXYLULOSE 5-PHOSPHATE
; FILE REFERENCE: 2037 US Divisional
; CURRENT APPLICATION NUMBER: US/10/046,583A
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: 09/626,589
; PRIOR FILING DATE: 2000-07-27
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 824
; TYPE: PRT
; FEATURE:
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: Residues 1-165 are from the thioredoxin sequence
; OTHER INFORMATION: found in the vector pET32 supplied by No. US20020168743Alagen.
; OTHER INFORMATION: Residues 166-824 represent the tDPS sequence from
; OTHER INFORMATION: Arabidopsis shown in SEQ ID NO:2.
US-10-046-583A-3

Query Match      11.6%; Score 50; DB 9; Length 824;
Best Local Similarity 100.0%; Pred. No. 3.2e-32;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHSSGLVPRGSGMKETAARKFERQHMDSPDLGTDDDDKAMADIGS 50
|||||
DB 116 MHHHHHSSGLVPRGSGMKETAARKFERQHMDSPDLGTDDDDKAMADIGS 165

RESULT 12
US-09-978-756-1
; Sequence 1, Application US/09978756
; Patent No. US20020160017A1
; GENERAL INFORMATION:
; APPLICANT: Holder, Anthony
; APPLICANT: Birdsell, Berry
```

```
; APPLICANT: Feeney, James
; APPLICANT: Morgan, William
; APPLICANT: Syed, Shabih
; TITLE OF INVENTION: Malaria Vaccine
; FILE REFERENCE: 18396/1005
; CURRENT APPLICATION NUMBER: US/09/978,756
; CURRENT FILING DATE: 2001-10-16
; PRIOR APPLICATION NUMBER: PCT/GB00/01558
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: 09/311,817
; PRIOR FILING DATE: 1999-05-13
; PRIOR APPLICATION NUMBER: 2,271,451
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: 9909072.2
; PRIOR FILING DATE: 1999-04-20
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 96
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-09-978-756-1

Query Match      10.7%; Score 46; DB 9; Length 96;
Best Local Similarity 100.0%; Pred. No. 8.8e-30;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

*QY 352 NSGCFRHLDERECKCLLNKQEGDKCVENPNPTCENNGGCCDADA 397
|||||
DB 15 NSGCFRHLDERECKCLLNKQEGDKCVENPNPTCENNGGCCDADA 60

RESULT 13
US-10-087-464-35
; Sequence 35, Application US/10087464
; Publication No. US20030059436A1
; GENERAL INFORMATION:
; APPLICANT: Chishti, Athar
; APPLICANT: Oh, Steven
; APPLICANT: Liu, David
; APPLICANT: Goel, Vikas
; APPLICANT: Li, Xuerong
; TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses There
; FILE REFERENCE: S1237/7019
; CURRENT APPLICATION NUMBER: US/10/087,464
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 06/272,930
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 35
; LENGTH: 114
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-10-087-464-35

Query Match      10.7%; Score 46; DB 9; Length 114;
Best Local Similarity 100.0%; Pred. No. 1e-29;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 NSGCFRHLDERECKCLLNKQEGDKCVENPNPTCENNGGCCDADA 397
|||||
DB 15 NSGCFRHLDERECKCLLNKQEGDKCVENPNPTCENNGGCCDADA 60

RESULT 14
US-09-681-938-1
; Sequence 1, Application US/09681938
; Publication No. US20030003584A1
; GENERAL INFORMATION:
; APPLICANT: Adams, Darius
; TITLE OF INVENTION: Liposomal Vector Binding Protein for Hepatocyte DNA Delivery
; FILE REFERENCE: 29147
```

; CURRENT APPLICATION NUMBER: US/09/681,938  
; PRIOR FILING DATE: 2001-06-28  
; PRIOR APPLICATION NUMBER: 09/681,938  
; PRIOR FILING DATE: 2001-06-28  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1  
; LENGTH: 201  
; TYPE: PRT  
; ORGANISM: Custom Liposomal Vector Binding Protein for Hepatocyte DNA Delivery  
; NAME/KEY: BINDING  
; LOCATION: (1)...(201)  
US-09-681-938-1

Query Match 10.7%; Score 46; DB 9; Length 201;  
Best Local Similarity 100.0%; Pred. No. 1.6e-29;  
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MHHHHSSGLVPRGSGMKETAARQHMDSPLGTDDDDKAMA 46  
|||||  
Db 1 MHHHHSSGLVPRGSGMKETAARQHMDSPLGTDDDDKAMA 46  
|||||

RESULT 15  
US-09-978-756-2  
; Sequence 2, Application US/09978756  
; Patent No. US20020160017A1  
; GENERAL INFORMATION:  
; APPLICANT: Birdsell, Anthony  
; APPLICANT: Feeney, James  
; APPLICANT: Morgan, William  
; APPLICANT: Syed, Shabih  
; TITLE OF INVENTION: Malaria Vaccine  
; FILE REFERENCE: 18396/1005  
; CURRENT APPLICATION NUMBER: US/09/978,756  
; CURRENT FILING DATE: 2001-10-16  
; PRIOR APPLICATION NUMBER: PCT/GB00/01558  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: 09/311,817  
; PRIOR FILING DATE: 1999-05-13  
; PRIOR APPLICATION NUMBER: 2,271,451  
; PRIOR FILING DATE: 1999-05-25  
; PRIOR APPLICATION NUMBER: 9909072.2  
; PRIOR FILING DATE: 1999-04-20  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 376  
; TYPE: PRT  
; ORGANISM: Plasmodium falciparum  
US-09-978-756-2

Query Match 10.7%; Score 46; DB 9; Length 376;  
Best Local Similarity 100.0%; Pred. No. 2.8e-29;  
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 352 NSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENNGGCDADA 397  
|||||  
Db 277 NSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENNGGCDADA 322  
|||||

RESULT 16  
US-10-087-464-34  
; Sequence 34, Application US/10087464  
; Publication No. US20030059436A1  
; GENERAL INFORMATION:  
; APPLICANT: Chishtli, Athar  
; APPLICANT: Oh, Steven  
; APPLICANT: Liu, David  
; APPLICANT: Goel, Vikas  
; APPLICANT: Li, Xuerong

; TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses There  
; FILE REFERENCE: SI237/7019  
; CURRENT APPLICATION NUMBER: US/10/087,464  
; CURRENT FILING DATE: 2002-03-01  
; PRIOR APPLICATION NUMBER: US 06/272,930  
; PRIOR FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 34  
; LENGTH: 376  
; TYPE: PRT  
; ORGANISM: Plasmodium falciparum  
US-10-087-464-34

Query Match 10.7%; Score 46; DB 9; Length 376;  
Best Local Similarity 100.0%; Pred. No. 2.8e-29;  
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 352 NSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENNGGCDADA 397  
|||||  
Db 277 NSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENNGGCDADA 322  
|||||

RESULT 17  
US-10-087-464-11  
; Sequence 11, Application US/10087464  
; Publication No. US20030059436A1  
; GENERAL INFORMATION:  
; APPLICANT: Chishtli, Athar  
; APPLICANT: Oh, Steven  
; APPLICANT: Liu, David  
; APPLICANT: Goel, Vikas  
; APPLICANT: Li, Xuerong  
; TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses There  
; FILE REFERENCE: SI237/7019  
; CURRENT APPLICATION NUMBER: US/10/087,464  
; CURRENT FILING DATE: 2002-03-01  
; PRIOR APPLICATION NUMBER: US 06/272,930  
; PRIOR FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 11  
; LENGTH: 378  
; TYPE: PRT  
; ORGANISM: Plasmodium falciparum  
US-10-087-464-11

Query Match 10.7%; Score 46; DB 9; Length 378;  
Best Local Similarity 100.0%; Pred. No. 2.8e-29;  
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 352 NSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENNGGCDADA 397  
|||||  
Db 279 NSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENNGGCDADA 324  
|||||

RESULT 18  
US-10-087-464-10  
; Sequence 10, Application US/10087464  
; Publication No. US20030059436A1  
; GENERAL INFORMATION:  
; APPLICANT: Chishtli, Athar  
; APPLICANT: Oh, Steven  
; APPLICANT: Liu, David  
; APPLICANT: Goel, Vikas  
; APPLICANT: Li, Xuerong  
; TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses There  
; FILE REFERENCE: SI237/7019  
; CURRENT APPLICATION NUMBER: US/10/087,464  
; CURRENT FILING DATE: 2002-03-01  
; PRIOR APPLICATION NUMBER: US 06/272,930  
; PRIOR FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 59

```
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 1639
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-10-087-464-10

Query Match      10.7%; Score 46; DB 9; Length 1639;
Best Local Similarity 100.0%; Pred. No. 9.4e-29;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 NSCCFRLHDERBCKLLNYKQBGDKCVENPNTCENNGGCCDADA 397
|||||
DB 1540 NSCCFRLHDERBCKLLNYKQBGDKCVENPNTCENNGGCCDADA 1585

RESULT 19
US-10-012-896-1011
; Sequence 1011, Application US/10012896
; Publication No. US20020183251A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; APPLICANT: Wantanabe, Yoshihiro
; APPLICANT: Meagher, Madeleine Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C27
; CURRENT APPLICATION NUMBER: US/10/012,896
; CURRENT FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 1011
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1011
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-896-1011

Query Match      10.4%; Score 45; DB.9; Length 355;
Best Local Similarity 100.0%; Pred. No. 1.7e-28;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHSSGLVPRGSGMKETAARKFERQHMDSPDLGTDDDDKAM 45
|||||
DB 116 MHHHHHSSGLVPRGSGMKETAARKFERQHMDSPDLGTDDDDKAM 160

RESULT 20
US-10-042-945-28
; Sequence 28, Application US/10042945
; Publication No. US20030045468A1
; GENERAL INFORMATION:
; APPLICANT: Fling, Steven P.
```

```
; APPLICANT: Foy, Teresa M.
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Wang, Aijun
; APPLICANT: Johnson, Jeffrey C.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Sutherland, R. Alec
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY,
; TITLE OF INVENTION: DIAGNOSIS AND MONITORING OF BREAST CANCER
; FILE REFERENCE: 210121.479C3
; CURRENT APPLICATION NUMBER: US/10/042,945
; CURRENT FILING DATE: 2002-01-08
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 55
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-042-945-28

Query Match      7.4%; Score 32; DB 9; Length 55;
Best Local Similarity 100.0%; Pred. No. 1e-18;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 GSGMKETAARKFERQHMDSPDLGTDDDDKAMA 46
|||||
DB 1 GSGMKETAARKFERQHMDSPDLGTDDDDKAMA 32

RESULT 21
US-09-757-417-28
; Sequence 28, Application US/09757417
; Patent No. US20020082216A1
; GENERAL INFORMATION:
; APPLICANT: Fanger, Gary R.
; APPLICANT: Foy, Theresa M.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: THERAPY, DIAGNOSIS AND MONITORING OF BREAST CANCER
; FILE REFERENCE: 210121.479C1
; CURRENT APPLICATION NUMBER: US/09/757,417
; CURRENT FILING DATE: 2001-01-08
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 55
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-757-417-28

Query Match      7.4%; Score 32; DB 10; Length 55;
Best Local Similarity 100.0%; Pred. No. 1e-18;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 GSGMKETAARKFERQHMDSPDLGTDDDDKAMA 46
|||||
DB 1 GSGMKETAARKFERQHMDSPDLGTDDDDKAMA 32

RESULT 22
US-09-872-712-1
; Sequence 1, Application US/09872712
; Publication No. US20030059461A1
; GENERAL INFORMATION:
; APPLICANT: Backer, Marina V.
; APPLICANT: Backer, Joseph M.
; TITLE OF INVENTION: MOLECULAR DELIVERY VEHICLE FOR DELIVERY
; TITLE OF INVENTION: OF SELECTED COMPOUNDS TO TARGETS
; FILE REFERENCE: 102131-200
; CURRENT APPLICATION NUMBER: US/09/872,712
; CURRENT FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: 60/209,660
; PRIOR FILING DATE: 2000-06-05
```

```

; SOFTWARE: PATENTCIN VER: 2.0
; SEQ ID NO 5
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial Sequence

```

RESULT 26  
US-09-096-749A-114  
; Sequence 114, Application US/09096749A  
; Patent No. US20020019517A1  
; GENERAL INFORMATION:  
; APPLICANT: Koieda, Shohei  
; TITLE OF INVENTION: ARTIFICIAL ANTIBODY POLYPEPTIDES  
; NUMBER OF SEQUENCES: 118  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth P.A.  
; STREET: 121 South Eighth Street, Ste. 1600  
; CITY: Minneapolis  
; STATE: MN  
; COUNTRY: USA  
; ZIP: 55402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0b  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/096.749A  
; FILING DATE: June 12, 1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ann S. Viksnins  
; REGISTRATION NUMBER: 37,748



REFERENCE/DOCKET NUMBER: 109.034US1  
TELEPHONE: (612) 373-6900  
TELEFAX: (612) 339-3061  
INFORMATION FOR SEQ ID NO: 114:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: Internal  
ORIGINAL SOURCE:  
US-09-096-749A-114

Query Match 3.5%; Score 15; DB 10; Length 20;  
Best Local Similarity 100.0%; Pred. No. 2.1e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HHHHHSSGLVPRGS 16  
Db 5 HHHHHSSGLVPRGS 19

## RESULT 27

US-10-060-275-5  
Sequence 5, Application US/10060275  
Publication No. US20030073828A1  
GENERAL INFORMATION:  
APPLICANT: DAI, ZIYU  
APPLICANT: SHI, LIFANG  
APPLICANT: HOOKER, BRIAN S.  
TITLE OF INVENTION: EPIMERASE GENE AND USE THEREOF  
FILE REFERENCE: 059440-0143  
CURRENT APPLICATION NUMBER: US/10/060,275  
CURRENT FILING DATE: 2002-06-11  
PRIOR APPLICATION NUMBER: 60/265,311  
PRIOR FILING DATE: 2001-02-01  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 5  
LENGTH: 21  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Synthetic amino  
OTHER INFORMATION: acid sequence

## US-10-060-275-5

Query Match 3.5%; Score 15; DB 9; Length 21;  
Best Local Similarity 100.0%; Pred. No. 2.2e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HHHHHSSGLVPRGS 16  
Db 5 HHHHHSSGLVPRGS 19

## RESULT 28

US-09-736-084-98  
Sequence 98, Application US/09736084  
Patent No. US20020107211A1  
GENERAL INFORMATION:

APPLICANT: THE ROCKEFELLER UNIVERSITY  
TITLE OF INVENTION: MODULATORS OF BODY WEIGHT, CORRESPONDING  
NUCLEIC ACIDS AND PROTEINS, AND DIAGNOSTIC AND THERAPEUTIC  
NUMBER OF SEQUENCES: 98  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack

STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
CURRENT APPLICATION DATA:  
PatentIn Release #1.0, Version #1.25  
APPLICATION NUMBER: US/09/736,084  
FILING DATE: 13-Dec-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/438,431  
FILING DATE: May 10, 1995  
APPLICATION NUMBER: 08/347,563  
FILING DATE: NO. US20020107211A1  
APPLICATION NUMBER: 08/292,345  
FILING DATE: August 17, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-087 CIP21  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521

INFORMATION FOR SEQ ID NO: 98:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
SEQUENCE DESCRIPTION: SEQ ID NO: 98:  
US-09-736-084-98

Query Match 3.5%; Score 15; DB 10; Length 21;  
Best Local Similarity 100.0%; Pred. No. 2.2e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HHHHHSSGLVPRGS 16  
Db 5 HHHHHSSGLVPRGS 19

## RESULT 29

US-09-736-084-12  
Sequence 12, Application US/09736084  
Patent No. US20020107211A1  
GENERAL INFORMATION:

APPLICANT: THE ROCKEFELLER UNIVERSITY  
TITLE OF INVENTION: MODULATORS OF BODY WEIGHT, CORRESPONDING  
NUCLEIC ACIDS AND PROTEINS, AND DIAGNOSTIC AND THERAPEUTIC

NUMBER OF SEQUENCES: 98

CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauber & Jackson

STREET: 411 Hackensack Avenue

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/736,084

FILING DATE: 13-Dec-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/438,431

; FILING DATE: May 10, 1995  
; APPLICATION NUMBER: 08/347,563  
; FILING DATE: NO. US20020107211A1ember 30, 1994  
; APPLICATION NUMBER: 08/292,345  
; FILING DATE: August 17, 1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jackson Esq., David A.  
; REGISTRATION NUMBER: 26,742  
; REFERENCE/DOCKET NUMBER: 600-1-087 CIP21  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201 487-5800  
; TELEFAX: 201 343-1684  
; TELEX: 133521  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 43 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 12:  
US-09-736-084-12

Query Match 3.5% Score 15; DB 10; Length 43;  
Best Local Similarity 100.0%; Pred. No. 3.9e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HHHHHSSGLVPRGS 16  
| | | | | | | | | | | | | | | | | |  
Db 5 HHHHHSSGLVPRGS 19

## RESULT 30

US-10-267-311-12  
; Sequence 12, Application US/10267311  
; Publication No. US20030050469A1  
; GENERAL INFORMATION:  
; APPLICANT: Siegel, Marvin  
; APPLICANT: Chu, N. Randall  
; APPLICANT: Mizzen, Lee A.  
; TITLE OF INVENTION: INDUCTION OF A THI-LIKE RESPONSE IN VITRO  
; FILE REFERENCE: 12071/002001  
; CURRENT APPLICATION NUMBER: US/10/267,311  
; CURRENT FILING DATE: 2002-10-09  
; PRIOR APPLICATION NUMBER: US/09/613,303  
; PRIOR FILING DATE: 2000-07-10  
; PRIOR APPLICATION NUMBER: US 60/143,757  
; PRIOR FILING DATE: 1999-07-08  
; NUMBER OF SEQ ID NOS: 55  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12  
; LENGTH: 121  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: fusion sequence  
US-10-267-311-12

Query Match 3.5% Score 15; DB 9; Length 121;  
Best Local Similarity 100.0%; Pred. No. 9.3e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HHHHHSSGLVPRGS 16  
| | | | | | | | | | | | | | | | | |  
Db 5 HHHHHSSGLVPRGS 19

## RESULT 31

US-09-981-286A-8  
; Sequence 8, Application US/09981286A  
; Publication No. US20020192799A1  
; GENERAL INFORMATION:  
; APPLICANT: Watowich, Stanley J.  
; APPLICANT: Weaver, Scott C.

; APPLICANT: Davey, Robert A.  
; TITLE OF INVENTION: Drug Discovery Methods  
; FILE REFERENCE: 265.00260101  
; CURRENT APPLICATION NUMBER: US/09/981,286A  
; CURRENT FILING DATE: 2001-10-15  
; PRIOR APPLICATION NUMBER: US 60/240,187  
; PRIOR FILING DATE: 2000-10-13  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 8  
; LENGTH: 124  
; TYPE: PRT  
; ORGANISM: BOS TAURUS  
US-09-981-286A-8

Query Match 3.5% Score 15; DB 9; Length 124;  
Best Local Similarity 100.0%; Pred. No. 9.5e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KETAAKFERQHMDs 33  
| | | | | | | | | | | | | | | | | |  
Db 1 KETAAKFERQHMDs 15

## RESULT 32

US-09-876-348A-19  
; Sequence 19, Application US/09876348A  
; Patent No. US20020172951A1  
; GENERAL INFORMATION:  
; APPLICANT: Horwath, K. L. and Myers, K. L.  
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding Type III Tenebrio  
; FILE REFERENCE: RB-125-RI  
; CURRENT APPLICATION NUMBER: US/09/876,348A  
; CURRENT FILING DATE: 2001-08-09  
; PRIOR APPLICATION NUMBER: 60/210,446  
; PRIOR FILING DATE: 2000-06-08  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: Microsoft Word  
; SEQ ID NO 19  
; LENGTH: 149  
; TYPE: PRT  
; ORGANISM: Tenebrio molitor  
; OTHER INFORMATION: Mature Protein with His-tag, Clone 2.2  
US-09-876-348A-19

Query Match 3.5% Score 15; DB 9; Length 149;  
Best Local Similarity 100.0%; Pred. No. 0.00011;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HHHHHSSGLVPRGS 16  
| | | | | | | | | | | | | | | | | |  
Db 5 HHHHHSSGLVPRGS 19

## RESULT 33

US-09-876-348A-23  
; Sequence 23, Application US/09876348A  
; Patent No. US20020172951A1  
; GENERAL INFORMATION:  
; APPLICANT: Horwath, K. L. and Myers, K. L.  
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding Type III Tenebrio  
; FILE REFERENCE: RB-125-RI  
; CURRENT APPLICATION NUMBER: US/09/876,348A  
; CURRENT FILING DATE: 2001-08-09  
; PRIOR APPLICATION NUMBER: 60/210,446  
; PRIOR FILING DATE: 2000-06-08  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: Microsoft Word  
; SEQ ID NO 23  
; LENGTH: 149  
; TYPE: PRT



; CURRENT APPLICATION NUMBER: US/09/876,796A  
; CURRENT FILING DATE: 2001-06-07  
; PRIOR APPLICATION NUMBER: 60/210,446  
; PRIOR FILING DATE: 2000-06-08  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: Microsoft Word  
; SEQ ID NO 19  
; LENGTH: 149  
; TYPE: PRT  
; ORGANISM: Tenebrio molitor  
; OTHER INFORMATION: Mature Protein with His-tag, Clone 2.2  
US-09-876-796A-19

Query Match 3.5%; Score 15; DB 9; Length 149;  
Best Local Similarity 100.0%; Pred. No. 0.00011;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HHHHHSSGLVPRGS 16  
|  
Db 5 HHHHHSSGLVPRGS 19

## RESULT 39

US-09-876-796A-23  
; Sequence 23, Application US/09876796A  
; Patent No. US20020173024A1  
; GENERAL INFORMATION:  
; APPLICANT: Horwath, K. L. and Easton, C. M.  
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding Type III Tenebrio  
; TITLE OF INVENTION: Antifreeze Proteins and Method for Assaying Activity.  
; FILE REFERENCE: RB-123-SEQ  
; CURRENT APPLICATION NUMBER: US/09/876,796A  
; CURRENT FILING DATE: 2001-06-07  
; PRIOR APPLICATION NUMBER: 60/210,446  
; PRIOR FILING DATE: 2000-06-08  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: Microsoft Word  
; SEQ ID NO 23  
; LENGTH: 149  
; TYPE: PRT  
; ORGANISM: Tenebrio molitor  
; OTHER INFORMATION: Mature Protein with His-tag, Clone 2.3  
US-09-876-796A-23

Query Match 3.5%; Score 15; DB 9; Length 149;  
Best Local Similarity 100.0%; Pred. No. 0.00011;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HHHHHSSGLVPRGS 16  
|  
Db 5 HHHHHSSGLVPRGS 19

## RESULT 40

US-09-876-796A-27  
; Sequence 27, Application US/09876796A  
; Patent No. US20020173024A1  
; GENERAL INFORMATION:  
; APPLICANT: Horwath, K. L. and Easton, C. M.  
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding Type III Tenebrio  
; TITLE OF INVENTION: Antifreeze Proteins and Method for Assaying Activity.  
; FILE REFERENCE: RB-123-SEQ  
; CURRENT APPLICATION NUMBER: US/09/876,796A  
; CURRENT FILING DATE: 2001-06-07  
; PRIOR APPLICATION NUMBER: 60/210,446  
; PRIOR FILING DATE: 2000-06-08  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: Microsoft Word  
; SEQ ID NO 27  
; LENGTH: 149  
; TYPE: PRT  
; ORGANISM: Tenebrio molitor  
; OTHER INFORMATION: Mature Protein with His-tag, Tm 13.17

## US-09-876-796A-27

Query Match 3.5%; Score 15; DB 9; Length 149;  
Best Local Similarity 100.0%; Pred. No. 0.00011;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HHHHHSSGLVPRGS 16  
|  
Db 5 HHHHHSSGLVPRGS 19

Search completed: May 12, 2003, 10:35:42  
Job time : 38 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

```
Run on:      May 12, 2003, 10:25:18 ; Search time 21 Seconds
              (without alignments)
              1773.045 Million cell
```

Title: US-10-057-531A-2  
 perfect score: 431  
 Sequence: 1 MHHHHHSSGLVPRGSGMKE.....TCECTKPDSPVLFDCIFCSS 431

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Listing first 1000 summaries

```
Database :      PIR_73:*
1:  pir1:*
2:  pir2:*
3:  pir3:*
4:  pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query			DB	ID	Description
		Match	Length				
1	231	53.6	1726	1	SAZQGM	major merozoite su	
2	190	44.1	1701	2	A54498	major merozoite su	
3	154	35.7	1726	2	A45948	major merozoite su	
4	145	33.6	651	2	A47282	merozoite surface	
5	135	31.3	1701	2	A26868	major merozoite su	
6	46	10.7	400	2	A45545	major merozoite su	
7	46	10.7	1639	2	S05603	major merozoite su	
8	24	5.6	1631	1	SAZQK1	major merozoite su	
9	23	5.3	1640	2	A24594	probable major sur	
10	15	3.5	124	1	NRBOB	pancreatic ribonuc	
11	15	3.5	124	1	NRANE	pancreatic ribonuc	
12	15	3.5	124	2	S08547	pancreatic ribonuc	
13	15	3.5	124	2	JC5560	pancreatic ribonuc	
14	15	3.5	150	1	NRBO	pancreatic ribonuc	
15	14	3.2	30	2	A05004	pancreatic ribonuc	
16	12	2.8	124	1	NRGN	pancreatic ribonuc	
17	12	2.8	124	1	NRSH	pancreatic ribonuc	
18	12	2.8	124	1	NRPH	pancreatic ribonuc	
19	12	2.8	124	1	NRPRH	pancreatic ribonuc	
20	12	2.8	124	2	S07141	pancreatic ribonuc	
21	12	2.8	125	4	A47498	seminal ribonuclea	
22	12	2.8	150	1	NRBOS	seminal ribonuclea	
23	11	2.6	124	1	NRDER	pancreatic ribonuc	
24	11	2.6	124	1	NRDEO	pancreatic ribonuc	
25	11	2.6	124	1	NRDEF	pancreatic ribonuc	
26	11	2.6	124	1	NRDEN	pancreatic ribonuc	
27	11	2.6	124	1	NRKEN	pancreatic ribonuc	
28	9	2.1	124	1	NRGF	pancreatic ribonuc	
29	9	2.1	124	1	NRGM	pancreatic ribonuc	

103	7	1.6	356	2	C91019	hypothetical prote	176	1.6	1019	2	T00117	dve protein - frui
104	7	1.6	358	2	E85863	hypothetical prote	177	1.6	1104	2	C72409	reverse gyrase - T
105	7	1.6	366	2	F81414	probable transmem	178	1.6	1122	2	T28130	hypothetical prote
106	7	1.6	373	2	B69860	conserved hypotet	179	1.6	1173	2	T30608	proteophosphoglyc
107	7	1.6	374	2	H88503	protein B0361.4 [i	180	1.6	1180	2	S69205	stripe a/b protein
108	7	1.6	382	2	C90407	conserved hypotet	181	1.6	1214	2	T47659	spliceosomal-like
109	7	1.6	385	2	S6284	hypothetical prote	182	1.6	1244	2	T19068	hypothetical prote
110	7	1.6	396	1	H70730	cytochrome P450 Rv	183	1.6	1306	2	S42659	acetyl-CoA carboxy
111	7	1.6	401	2	S53405	probable membrane	184	1.6	1308	2	T15280	hypothetical prote
112	7	1.6	405	2	E90154	primase (Amino-end	185	1.6	1339	2	A55301	1,3-beta-D-glucan-
113	7	1.6	418	2	G84546	probable tetracycl	186	1.6	1441	2	T13717	CRAG protein - fru
114	7	1.6	420	2	T39712	hypothetical prote	187	1.6	1446	2	T30916	carboxypeptidase D
115	7	1.6	427	2	T42516	hypothetical prote	188	1.6	1597	1	BVFFSL	sol protein, large
116	7	1.6	435	2	A31596	clathrin coat asse	189	1.6	1597	2	T08428	gene small optic l
117	7	1.6	435	2	T49327	assembly protein 5	190	1.6	1625	2	T02921	acetyl-CoA carboxy
118	7	1.6	435	2	G02088	shikimate transpor	191	1.6	1685	2	T02750	acetyl-CoA carboxy
119	7	1.6	438	2	G64962	shikimate transpor	192	1.6	1726	2	A39401	major blood-stage
120	7	1.6	438	2	B90976	shikimate transpor	193	1.6	1751	2	A45604	acetyl-CoA carboxy
121	7	1.6	438	2	H85822	probable transport	194	1.6	2311	2	T06161	acetyl-CoA carboxy
122	7	1.6	459	2	T45576	anthranilate N-hyd	195	1.6	2325	2	T02235	his operon leader
123	7	1.6	466	2	E75201	pyridoxal phosphat	196	1.6	16	1	LFEC	his operon leader
124	7	1.6	473	2	T48985	hypothetical prote	197	1.6	16	1	C90981	his operon leader
125	7	1.6	476	2	S57963	methyl CpG binding	198	1.6	16	1	A85827	ig heavy chain CDR
126	7	1.6	490	2	S52830	HMS1 protein - yea	199	1.6	21	2	T02227	hypothetical prote
127	7	1.6	500	1	E75342	IMP dehydrogenase	200	1.6	32	2	B82378	hypothetical prote
128	7	1.6	500	2	A55568	monocarboxylate tr	201	1.6	52	2	C90235	H repeat-associat
129	7	1.6	503	2	H87596	tryptophan halogen	202	1.6	54	2	A99983	degenerate transpo
130	7	1.6	513	2	D64495	hypothetical prote	203	1.6	55	2	H97965	hypothetical prote
131	7	1.6	513	2	S39691	urp-hexose-1-phosp	204	1.6	56	2	A12134	histidine and glut
132	7	1.6	513	2	B64431	nodulation factor	205	1.6	57	2	H64698	probable coiled-co
133	7	1.6	515	1	T38946	phosphoprotein pho	206	1.6	60	2	F81308	hypothetical prote
134	7	1.6	528	2	B75364	extracellular solu	207	1.6	60	2	D84411	hypothetical prote
135	7	1.6	532	2	T15354	hypothetical prote	208	1.6	64	2	AC2576	hypothetical prote
136	7	1.6	548	2	A05032	rpcC protein homol	209	1.6	64	2	S57787	hypothetical prote
137	7	1.6	554	2	JE0303	propanediol dehydr	210	1.6	70	2	E96930	uncharacterized Fe
138	7	1.6	554	2	A56111	glycerol dehydrata	211	1.6	76	1	B64700	carbon storage reg
139	7	1.6	554	2	AC0760	hypothetical prote	212	1.6	76	2	E71820	probable carbon st
140	7	1.6	555	2	T32105	hypothetical prote	213	1.6	76	2	A97869	hypothetical prote
141	7	1.6	590	2	T49672	related to a-agglu	214	1.6	77	2	C70306	conserved hypotet
142	7	1.6	623	1	S33167	gene pointed prote	215	1.6	77	2	D11821	probable histidine
143	7	1.6	624	2	H90071	hypothetical prote	216	1.6	77	2	T16436	hypothetical prote
144	7	1.6	642	2	A81798	RNA polymerase sig	217	1.6	77	2	E97137	hypothetical prote
145	7	1.6	642	2	F81072	RNA polymerase sig	218	1.6	78	2	AB1162	flagellar switch p
146	7	1.6	659	2	T20753	hypothetical prote	219	1.6	80	2	C69204	weakly flagellar s
147	7	1.6	666	2	T17396	vrp protein - Dic	220	1.6	82	2	T01133	hypothetical prote
148	7	1.6	682	2	A44493	serum-inducible ki	221	1.6	83	2	I46058	hypothetical prote
149	7	1.6	684	2	H96918	probable regulator	222	1.6	83	2	E72151	caldesmon - bovine
150	7	1.6	746	2	T28004	hypothetical prote	223	1.6	84	2	S33432	B13R protein - var
151	7	1.6	753	2	S48267	probable membrane	224	1.6	84	2	C97194	hypothetical prote
152	7	1.6	768	2	AB1085	hypothetical prote	225	1.6	86	2	B26393	endonexin - bovine
153	7	1.6	792	2	B82752	penicillin binding	226	1.6	86	2	T27405	hypothetical prote
154	7	1.6	818	2	JC4058	fibroblast growth	227	1.6	87	2	G83523	conserved hypotet
155	7	1.6	825	1	A60386	interleukin-4 rece	228	1.6	88	2	B46264	thioredoxin 2 - sl
156	7	1.6	829	2	JC4583	fibroblast growth	229	1.6	90	2	S78047	DNA-directed RNA p
157	7	1.6	846	2	C82135	chitinase VC1952 [	230	1.6	91	2	F70142	glu-tRNA amidotran
158	7	1.6	869	2	A55384	transcription fact	231	1.6	92	2	T34146	hypothetical prote
159	7	1.6	876	2	D96558	probable protein k	232	1.6	93	2	F84585	hypothetical prote
160	7	1.6	885	2	G95980	probable nitrate r	233	1.6	93	2	T48043	hypothetical prote
161	7	1.6	896	2	T51891	hypothetical prote	234	1.6	93	2	H97840	hypothetical prote
162	7	1.6	897	2	C90561	hypothetical prote	235	1.6	95	2	F71731	glutaredoxin 3 (gr
163	7	1.6	913	2	D82885	multiple banded an	236	1.6	96	1	W4LE35	Ed protein - human
164	7	1.6	925	2	T16235	hypothetical prote	237	1.6	101	2	F90262	hypothetical prote
165	7	1.6	946	2	T16297	hypothetical prote	238	1.6	101	2	B44971	hypothetical prote
166	7	1.6	951	2	S52728	H+-exporting ATPas	239	1.6	102	2	S09892	hypothetical prote
167	7	1.6	956	2	A45506	H+-exporting ATPas	240	1.6	102	2	T30119	hypothetical prote
168	7	1.6	956	2	A43637	H+-exporting ATPas	241	1.6	102	2	C97733	glutaredoxin 3 lim
169	7	1.6	956	2	S50751	H+-exporting ATPas	242	1.6	103	2	C84448	hypothetical prote
170	7	1.6	956	2	T52412	H+-exporting ATPas	243	1.6	105	2	B97700	thioredoxin (impor
171	7	1.6	957	2	T03846	probable plasma me	244	1.6	105	2	T49655	thioredoxin
172	7	1.6	960	2	T06688	H+-exporting ATPas	245	1.6	106	2	S33357	thioredoxin - Stre
173	7	1.6	963	2	T12087	H+-exporting ATPas	246	1.6	107	2	S55137	hypothetical prote
174	7	1.6	989	2	T46183	zinc proteinase (E	247	1.6	110	2	S55208	hypothetical prote
175	7	1.6	1001	2	T28997	hypothetical prote	248	1.6	110	2	S55208	hypothetical prote

249	6	1.4	110	2	A71454	hypothetical prote	322	6	1.4	143	2	JC7223	lebecin-like prote
250	6	1.4	110	2	T17639	hypothetical prote	323	6	1.4	143	2	A9315	conserved hypotet
251	6	1.4	111	2	T16516	hypothetical prote	324	6	1.4	144	2	S44547	hypothetical prote
252	6	1.4	113	2	S08455	hypothetical prote	325	6	1.4	144	2	A84188	hypothetical prote
253	6	1.4	113	2	C81220	hypothetical prote	326	6	1.4	144	2	G64423	phospholipase A2 (
254	6	1.4	113	2	G81990	hypothetical prote	327	6	1.4	146	1	PSFGA	YU8H12.12 (import
255	6	1.4	114	2	S37150	asr2 protein - tom	328	6	1.4	146	2	C86187	hypothetical prote
256	6	1.4	114	2	T20095	hypothetical prote	329	6	1.4	147	2	E82923	hypothetical prote
257	6	1.4	115	2	H72583	hypothetical prote	330	6	1.4	149	2	G64414	hypothetical prote
258	6	1.4	116	2	G64433	hypothetical prote	331	6	1.4	150	2	S54517	hypothetical prote
259	6	1.4	116	2	G84032	hypothetical prote	332	6	1.4	150	2	E83057	hypothetical prote
260	6	1.4	117	2	G64335	hypothetical prote	333	6	1.4	150	2	S74649	hypothetical prote
261	6	1.4	119	2	A86605	iojap superfamily	334	6	1.4	150	2	C64319	hypothetical prote
262	6	1.4	119	2	B72020	iojap homolog - Ch	335	6	1.4	152	2	G72213	ferric uptake regu
263	6	1.4	119	2	A11591	B. subtilis PBSX p	336	6	1.4	152	2	T51796	hypothetical prote
264	6	1.4	121	1	A36270	hemoglobin - Tetra	337	6	1.4	153	2	I53530	pancreatic ribonuc
265	6	1.4	122	2	F83167	hypothetical prote	338	6	1.4	153	2	A82687	conserved hypotet
266	6	1.4	122	2	T01558	auxin-induced prot	339	6	1.4	153	2	D97469	ynaf protein (AE00
267	6	1.4	123	2	B31596	pancreatic ribonuc	340	6	1.4	153	2	C97218	conserved membrane
268	6	1.4	124	1	NRWHK	clathrin coat asse	341	6	1.4	154	2	E82457	hypothetical prote
269	6	1.4	124	1	NRCEB	pancreatic ribonuc	342	6	1.4	154	2	E90588	conserved hypotet
270	6	1.4	124	1	NRGPA	pancreatic ribonuc	343	6	1.4	154	2	S40522	Huntington's disea
271	6	1.4	124	1	NRUI	pancreatic ribonuc	344	6	1.4	155	2	C86637	hypothetical prote
272	6	1.4	124	2	C70442	ribosomal protein	345	6	1.4	156	1	NRH01	pancreatic ribonuc
273	6	1.4	126	2	H96914	chemotaxis respons	346	6	1.4	156	2	B97836	bacterioferritin c
274	6	1.4	126	2	T03757	hypothetical prote	347	6	1.4	156	2	A98073	hypothetical prote
275	6	1.4	128	1	NROW2	pancreatic ribonuc	348	6	1.4	156	2	A87567	conserved hypotet
276	6	1.4	128	2	A33083	pancreatic ribonuc	349	6	1.4	157	2	AD2282	hypothetical prote
277	6	1.4	128	2	T15017	hypothetical prote	350	6	1.4	158	2	F69741	gltx 5'-region con
278	6	1.4	128	2	G81220	hypothetical prote	351	6	1.4	159	2	G90103	rubredoxin (import
279	6	1.4	130	2	D17107	thioredoxin (trxa)	352	6	1.4	161	2	E75043	conserved hypotet
280	6	1.4	130	2	T29887	hypothetical prote	353	6	1.4	161	2	F71117	conserved hypotet
281	6	1.4	130	2	S14983	extensin class I (	354	6	1.4	161	2	S61624	probable membrane
282	6	1.4	131	2	A97791	nifu protein (limp	355	6	1.4	162	2	S05712	phycocyanin 3 alph
283	6	1.4	131	2	H95339	hypothetical prote	356	6	1.4	162	2	H83792	hypothetical prote
284	6	1.4	132	2	S66476	cytochrome P450 (C	357	6	1.4	162	2	F95165	hypothetical prote
285	6	1.4	132	2	I39193	gene HOXA1 protein	358	6	1.4	162	2	E98031	hypothetical prote
286	6	1.4	133	2	AB2421	hypothetical prote	359	6	1.4	163	1	F81938	cytochrome c552 NM
287	6	1.4	133	2	B30242	stem cell protein	360	6	1.4	163	2	D71169	hypothetical prote
288	6	1.4	135	2	S69464	hypothetical prote	361	6	1.4	164	1	TVBE11	transforming prote
289	6	1.4	135	2	H87008	conserved hypotet	362	6	1.4	164	2	S29550	rhizopuspepsin (PC
290	6	1.4	135	2	T49640	hypothetical prote	363	6	1.4	165	2	A81382	shikimate kinase (
291	6	1.4	136	1	GGICE3	globin CTF-III - m	364	6	1.4	165	2	S37679	hypothetical prote
292	6	1.4	136	1	GGIC3	hemoglobin III - m	365	6	1.4	166	2	S37738	hypothetical 18.6K
293	6	1.4	136	2	T22959	hypothetical prote	366	6	1.4	166	2	T31015	hypothetical prote
294	6	1.4	137	1	T44983	methylmalonyl-CoA	367	6	1.4	166	2	T39825	hypothetical prote
295	6	1.4	137	2	E72465	hypothetical prote	368	6	1.4	167	1	ORECCW	purine binding che
296	6	1.4	137	2	B69039	hypothetical prote	369	6	1.4	167	2	S20065	pancreatic-type ri
297	6	1.4	137	2	AF0087	flagellar switch p	370	6	1.4	167	2	A26143	chemotaxis protein
298	6	1.4	137	2	AE1277	E. coli MutT prote	371	6	1.4	167	2	E90953	positive regulator
299	6	1.4	137	2	AE1640	cytochrome c552 NM	372	6	1.4	167	2	A85802	purine binding che
300	6	1.4	138	1	B81167	hypothetical prote	373	6	1.4	167	2	AD0746	hypothetical prote
301	6	1.4	138	2	AE1968	hypothetical prote	374	6	1.4	167	2	D70385	hypothetical prote
302	6	1.4	138	2	G69232	hypothetical prote	375	6	1.4	168	2	A95183	dihydrofolate redu
303	6	1.4	138	2	E90531	hypothetical prote	376	6	1.4	168	2	D98050	signal peptidase I
304	6	1.4	139	2	T33968	hypothetical prote	377	6	1.4	168	2	G70132	hypothetical prote
305	6	1.4	140	2	T06554	probable profilin	378	6	1.4	168	2	AG2257	probable GDP-L-fuc
306	6	1.4	140	2	A54523	histidine-rich pro	379	6	1.4	169	2	E90983	GDP-mannose mannos
307	6	1.4	140	2	B86683	prophage p11 prote	380	6	1.4	169	2	H85828	CT144 hypotetrical
308	6	1.4	140	2	C86800	prophage p13 prote	381	6	1.4	169	2	E86522	conserved hypotet
309	6	1.4	140	2	T19083	hypothetical prote	382	6	1.4	169	2	G81570	hypothetical prote
310	6	1.4	141	2	T15210	probable cytochrom	383	6	1.4	170	2	T31938	gene l15 protein -
311	6	1.4	141	2	A11105	ribosomal protein	384	6	1.4	170	2	T03322	hypothetical prote
312	6	1.4	141	2	A11467	ribosomal protein	385	6	1.4	171	2	A99415	hypothetical prote
313	6	1.4	141	2	T06553	probable profilin	386	6	1.4	171	2	G90532	hypothetical prote
314	6	1.4	141	2	B98145	hypothetical prote	387	6	1.4	172	2	T51065	hypothetical prote
315	6	1.4	142	2	AB3143	hypothetical prote	388	6	1.4	172	2	T34464	hypothetical prote
316	6	1.4	142	2	S28693	hypothetical prote	389	6	1.4	173	2	E64540	hypothetical prote
317	6	1.4	143	1	B64421	conserved hypotet	390	6	1.4	174	2	S01189	NADH2 dehydrogenas
318	6	1.4	143	2	T12144	hypothetical prote	391	6	1.4	174	2	B30020	NADH2 dehydrogenas
319	6	1.4	143	2	F86168	hypothetical prote	392	6	1.4	174	2	T02869	globulin.1 precurs
320	6	1.4	143	2	T43180	hypothetical prote	393	6	1.4	174	2	S38258	mannose-binding ie
321	6	1.4	143	2	T40426	hypothetical prote	394	6	1.4	174	2	F75606	hypothetical prote

395	6	1.4	175	2	JE0189	inorganic diphosph	468	6	1.4	206	2	H98265	hypothetical prote
396	6	1.4	175	2	F81437	formylmethionine d	469	6	1.4	207	2	H85070	hypothetical prote
397	6	1.4	175	2	C39141	transcription repr	470	6	1.4	208	2	F64244	ATP synthase B cha
398	6	1.4	175	2	S09784	hypothetical prote	471	6	1.4	208	2	S55930	het-c4 protein - p
399	6	1.4	175	2	S75258	hypothetical prote	472	6	1.4	208	2	D96543	unknown protein [i
400	6	1.4	176	2	T48265	hypothetical prote	473	6	1.4	208	2	T24446	hypothetical prote
401	6	1.4	176	2	G89977	conserved hypotet	474	6	1.4	210	1	STONC	conserved hypotet
402	6	1.4	177	2	F81175	inorganic pyrophos	475	6	1.4	210	2	T41553	thymidylate kinase
403	6	1.4	177	2	G81930	probable inorganic	476	6	1.4	210	2	S28955	dTMP kinase (EC 2.
404	6	1.4	178	2	D97160	uracil phosphoribo	477	6	1.4	210	2	S15186	somatotropin - chi
405	6	1.4	178	2	T08444	hypothetical prote	478	6	1.4	210	2	S50118	somatotropin precu
406	6	1.4	179	2	AD2427	hypothetical prote	479	6	1.4	210	2	JS0179	somatotropin precu
407	6	1.4	180	2	S08624	hydrogenase (EC 1.	480	6	1.4	210	2	S03709	somatotropin precu
408	6	1.4	180	2	H91075	formate hydrogenly	481	6	1.4	210	2	S06489	somatotropin II pr
409	6	1.4	180	2	G85920	formate hydrogenly	482	6	1.4	210	2	A31363	somatotropin precu
410	6	1.4	180	2	AD0846	formate hydrogenly	483	6	1.4	210	2	A23154	somatotropin I pre
411	6	1.4	180	2	H69115	probable orotate p	484	6	1.4	210	2	S15140	somatotropin precu
412	6	1.4	180	2	F84507	hypothetical prote	485	6	1.4	210	2	S08389	hypothetical prote
413	6	1.4	180	2	E84774	probable RING zinc	486	6	1.4	210	2	S27489	probable lipophorbo
414	6	1.4	180	2	F83252	hypothetical prote	487	6	1.4	210	2	F90404	purine phosphoribo
415	6	1.4	181	2	G65024	Hydrogenase-4 comp	488	6	1.4	210	2	S77469	hypothetical prote
416	6	1.4	181	2	F91047	hydrogenase 4 Fe-S	489	6	1.4	211	2	B82901	hypothetical prote
417	6	1.4	181	2	B85892	hydrogenase 4 Fe-S	490	6	1.4	212	2	D72235	conserved hypotet
418	6	1.4	181	2	B84488	Achilia retroelmen	491	6	1.4	212	2	B82920	hypothetical prote
419	6	1.4	182	2	S19203	gene HOX2.8 protei	492	6	1.4	214	2	D90185	SSU ribosomal prot
420	6	1.4	182	2	AC1653	Listeria prophage	493	6	1.4	214	2	B70470	hypothetical prote
421	6	1.4	183	2	F71641	ATP synthase delta	494	6	1.4	215	1	WMBE21	UL14 protein - hum
422	6	1.4	183	2	T14653	hypothetical prote	495	6	1.4	215	2	S24717	pyroglutaryl-pepti
423	6	1.4	183	2	E69353	hypothetical prote	496	6	1.4	216	2	AB2582	glutamine amidotra
424	6	1.4	184	2	D87605	transcription regu	497	6	1.4	216	2	H69553	conserved hypotet
425	6	1.4	184	2	B72601	hypothetical prote	498	6	1.4	217	2	T37859	probable transcrip
426	6	1.4	185	2	JK0021	somatotropin - ski	499	6	1.4	217	2	T31940	hypothetical prote
427	6	1.4	185	2	T31939	hypothetical prote	500	6	1.4	217	2	B71203	hypothetical prote
428	6	1.4	185	2	D85040	hypothetical prote	501	6	1.4	217	2	S27829	gametocytogenesis
429	6	1.4	185	2	D71854	fkbp-type peptidyl	502	6	1.4	218	2	S58769	brain-specific pro
430	6	1.4	185	2	C64660	peptidyl-prolyl ci	503	6	1.4	218	2	C82828	hypothetical prote
431	6	1.4	187	2	G75273	conserved hypotet	504	6	1.4	219	2	S66947	hypothetical prote
432	6	1.4	187	2	H82933	hypothetical prote	505	6	1.4	220	2	B81333	probable membrane
433	6	1.4	188	2	A32840	hypothetical 22k p	506	6	1.4	221	2	H97363	amidotransferase h
434	6	1.4	189	2	T36736	probable secreted	507	6	1.4	221	2	B81253	probable peptide A
435	6	1.4	189	2	F71978	hypothetical prote	508	6	1.4	221	2	JC4761	recombination acti
436	6	1.4	189	2	E64527	hypothetical prote	509	6	1.4	222	2	H70192	endonuclease III (
437	6	1.4	189	2	AE1446	hypothetical prote	510	6	1.4	222	2	I39192	gene HOXA1 protein
438	6	1.4	190	2	B81386	probable scaffold	511	6	1.4	222	2	H82530	hypothetical prote
439	6	1.4	191	2	A64704	hypothetical prote	512	6	1.4	222	2	E97538	phosphoglycolate p
440	6	1.4	191	2	A71814	hypothetical prote	513	6	1.4	222	2	AH2757	hydrolase [importe
441	6	1.4	191	2	E75132	molybdopterlin-gua	514	6	1.4	224	2	S61386	icmy protein - Leg
442	6	1.4	192	2	G71516	endopeptidase Clp	515	6	1.4	225	2	AG2427	hypothetical prote
443	6	1.4	192	2	D83273	conserved hypotet	516	6	1.4	225	2	G75448	conserved hypotet
444	6	1.4	192	2	T39367	hypothetical prote	517	6	1.4	225	2	F86498	pts IIA protein wi
445	6	1.4	193	2	D69212	conserved hypotet	518	6	1.4	225	2	A84786	probable cis-Golgi
446	6	1.4	195	2	T14965	hypothetical prote	519	6	1.4	225	2	G71533	probable pts IIA p
447	6	1.4	196	2	PD0004	self-incompatibili	520	6	1.4	225	2	F72124	pts IIA protein +
448	6	1.4	196	2	B81672	endopeptidase Clp	521	6	1.4	226	2	A32302	chemotaxis protein
449	6	1.4	196	2	S41271	ribosomal protein	522	6	1.4	227	2	B90400	hypothetical prote
450	6	1.4	196	2	S41263	ribosomal protein	523	6	1.4	228	2	AE2200	two-component resp
451	6	1.4	196	2	S41262	ribosomal protein	524	6	1.4	228	2	S51738	translin - human
452	6	1.4	196	2	S41257	ribosomal protein	525	6	1.4	228	2	F90130	hypothetical prote
453	6	1.4	196	2	I40130	outer surface prot	526	6	1.4	229	2	S42376	hypothetical prote
454	6	1.4	196	2	AI1361	probable scaffold	527	6	1.4	230	2	B90525	protoporphirogen o
455	6	1.4	199	1	G64070	imidazoleglycerol-	528	6	1.4	230	2	C87200	conserved hypotet
456	6	1.4	199	2	AD2591	somatotropin precu	529	6	1.4	230	2	T47866	regulatory protein
457	6	1.4	199	2	H88108	protein C46E10.8 l	530	6	1.4	231	1	A32667	NAD(P)H2 dehydroge
458	6	1.4	200	1	A37146	ribosomal protein	531	6	1.4	231	1	TRPCTR	trypsin (EC 3.4.21
459	6	1.4	200	2	B35116	anthranilate synth	532	6	1.4	231	2	T06258	superoxide dismuta
460	6	1.4	200	2	C83519	anthranilate synth	533	6	1.4	231	2	T06801	probable superoxid
461	6	1.4	201	2	AC3509	dephospho-CoA kina	534	6	1.4	231	2	S41570	chalcone isomerase
462	6	1.4	201	2	I40550	hypothetical prote	535	6	1.4	231	2	D84934	50S ribosomal prot
463	6	1.4	202	2	E75053	hypothetical prote	536	6	1.4	231	2	A75103	phosphoglycolate p
464	6	1.4	204	2	G83060	probable ribosomal	537	6	1.4	232	2	B64442	hypothetical prote
465	6	1.4	205	2	E97135	ribonuclease D [im	538	6	1.4	232	2	T17553	hypothetical prote
466	6	1.4	206	2	T25384	hypothetical prote	539	6	1.4	233	2	T25295	hypothetical prote
467	6	1.4	206	2	AI3018	conserved hypotet	540	6	1.4	233	2	AB1366	protein gp33 [Bact



541	6	1.4	233	2	A69412	hypothetical prote	614	6	1.4	261	1	S00812	glucose 1-dehydrog
542	6	1.4	234	2	D84075	hypothetical prote	615	6	1.4	261	2	A33528	glucose 1-dehydrog
543	6	1.4	235	2	D89101	protein F25E5.8 [i	616	6	1.4	261	2	JS0385	glucose 1-dehydrog
544	6	1.4	236	2	T24882	hypothetical prote	617	6	1.4	261	2	I39853	glucose 1-dehydrog
545	6	1.4	237	2	S41512	Brn-3b protein - m	618	6	1.4	261	2	I40225	glucose 1-dehydrog
546	6	1.4	237	2	D83969	RNA polymerase spo	619	6	1.4	261	2	I40225	glucose 1-dehydrog
547	6	1.4	237	2	S73263	hypothetical prote	620	6	1.4	261	2	D69629	glucose 1-dehydrog
548	6	1.4	237	2	H75374	beta-phosphoglucom	621	6	1.4	261	2	A70666	probable modA prot
549	6	1.4	237	2	A71082	hypothetical prote	622	6	1.4	261	2	B42882	motility protein (
550	6	1.4	237	2	E90555	hypothetical prote	623	6	1.4	262	2	S02299	glucose 1-dehydrog
551	6	1.4	237	2	S28560	early light-induce	624	6	1.4	262	2	F84994	glutamate racemase
552	6	1.4	237	2	G75076	hypothetical prote	625	6	1.4	262	2	I64200	hypothetical prote
553	6	1.4	238	2	F89854	hypothetical prote	626	6	1.4	262	2	S44803	hypothetical prote
554	6	1.4	238	2	C37416	bloodstream-specif	627	6	1.4	263	2	S01227	glucose 1-dehydrog
555	6	1.4	239	1	JU0083	transcription init	628	6	1.4	263	2	G84547	hypothetical prote
556	6	1.4	239	1	B39441	transcription init	629	6	1.4	263	2	H86147	glucose 1-dehydrog
557	6	1.4	239	2	T51270	hypothetical prote	630	6	1.4	264	2	A81362	TiN6.7 protein - A
558	6	1.4	239	2	E86346	Fl8F4.1 protein -	631	6	1.4	264	2	AE1362	probable terminase
559	6	1.4	239	2	A13117	hypothetical prote	632	6	1.4	265	2	D81315	hypothetical prote
560	6	1.4	240	2	S66275	proline-rich prote	633	6	1.4	265	2	T42419	hypothetical prote
561	6	1.4	240	2	G71968	hypothetical prote	634	6	1.4	265	2	T40259	hypothetical prote
562	6	1.4	240	2	B64540	conserved hypotet	635	6	1.4	266	2	C81254	hypothetical prote
563	6	1.4	240	2	AC1735	protein gp33 (Bact	636	6	1.4	266	2	E69746	conserved hypotet
564	6	1.4	240	2	F98065	hypothetical prote	637	6	1.4	267	2	B82541	ABC transporter pe
565	6	1.4	240	2	F87537	tomB protein, prob	638	6	1.4	267	2	D84750	hypothetical prote
566	6	1.4	240	2	T45727	hypothetical prote	639	6	1.4	268	2	A58446	Ig heavy chain v r
567	6	1.4	241	2	H84864	hypothetical prote	640	6	1.4	269	2	D82333	hypothetical prote
568	6	1.4	241	2	D64498	hypothetical prote	641	6	1.4	270	2	A26480	conserved hypotet
569	6	1.4	241	2	T22075	hypothetical prote	642	6	1.4	270	2	T41759	knob protein - mal
570	6	1.4	242	2	T31950	hypothetical prote	643	6	1.4	271	2	H85774	LEF-1 or fl14 - Bomb
571	6	1.4	243	2	B29624	spherulin lb precu	644	6	1.4	271	2	H85774	hypothetical prote
572	6	1.4	244	2	S73022	probable membrane	645	6	1.4	272	2	T26640	hypothetical prote
573	6	1.4	244	2	T08438	hypothetical prote	646	6	1.4	272	2	H70335	hypothetical prote
574	6	1.4	244	2	A89752	protein C33E10.1 [	647	6	1.4	274	1	A57691	hypothetical prote
575	6	1.4	245	2	H96932	hypothetical prote	648	6	1.4	274	2	T22562	NAD(P)H2 dehydrog
576	6	1.4	246	2	T38833	hypothetical prote	649	6	1.4	274	2	D91178	probable APAC-type
577	6	1.4	246	2	C82541	hypothetical UPF00	650	6	1.4	274	2	E86024	probable ARAC-type
578	6	1.4	247	1	TRDG	ABC transporter Ar	651	6	1.4	274	2	S47736	probable transcript
579	6	1.4	247	2	C69761	trypsin (EC 3.4.21	652	6	1.4	274	2	A89816	hypothetical prote
580	6	1.4	247	2	G82202	ABC-type transpor	653	6	1.4	274	2	H70900	hypothetical prote
581	6	1.4	247	2	E86250	hypothetical prote	654	6	1.4	275	2	T02334	hypothetical prote
582	6	1.4	248	1	B64462	protein F25C20.16	655	6	1.4	275	2	G91011	probable urease ac
583	6	1.4	248	2	T38025	dethiobiotin synth	656	6	1.4	275	2	A85856	probable elongatio
584	6	1.4	248	2	T33230	probable vacuolar	657	6	1.4	275	2	B64986	hypothetical 30.9
585	6	1.4	248	2	T22755	hypothetical prote	658	6	1.4	275	2	AH1924	hypothetical prote
586	6	1.4	248	2	B90345	hypothetical prote	659	6	1.4	276	2	T51437	hypothetical prote
587	6	1.4	248	2	F97414	hypothetical prote	660	6	1.4	276	2	B87203	probable antibioti
588	6	1.4	248	2	AB2632	hypothetical prote	661	6	1.4	277	2	A64516	hypothetical prote
589	6	1.4	248	2	T26412	hypothetical prote	662	6	1.4	277	2	S22494	rRNA N-glycosidase
590	6	1.4	249	2	T51687	myb-related transc	663	6	1.4	277	2	A37416	bloodstream-specif
591	6	1.4	249	2	C90526	conserved hypotet	664	6	1.4	277	2	T20065	hypothetical prote
592	6	1.4	251	2	E70521	probable phosphotr	665	6	1.4	278	2	T31942	hypothetical prote
593	6	1.4	251	2	A64456	hypothetical prote	666	6	1.4	278	2	T31935	hypothetical prote
594	6	1.4	253	2	T04642	hypothetical prote	667	6	1.4	279	2	S58402	hypothetical prote
595	6	1.4	253	2	S76719	hypothetical prote	668	6	1.4	279	2	S8402	synaptotagmin V -
596	6	1.4	253	2	T18978	hypothetical prote	669	6	1.4	279	2	D82281	ferric vibriobactl
597	6	1.4	253	2	F72710	hypothetical prote	670	6	1.4	279	2	G95117	Hemk protein (limp
598	6	1.4	254	2	AC1118	triosephosphate is	671	6	1.4	280	2	E97987	hypothetical prote
599	6	1.4	254	2	AE1478	triosephosphate is	672	6	1.4	280	2	T15715	hypothetical prote
600	6	1.4	254	2	S68481	hypothetical prote	673	6	1.4	280	2	B36144	unoporphyrinogen m
601	6	1.4	255	2	A41511	Brn-3a protein - m	674	6	1.4	280	2	S25326	hypothetical prote
602	6	1.4	255	2	B71490	hypothetical prote	675	6	1.4	280	2	D97100	uncharacterized pr
603	6	1.4	256	2	A32017	beta-lactamase (EC	676	6	1.4	281	2	E70173	dimethyladenosine
604	6	1.4	256	2	H75404	outer membrane pro	677	6	1.4	281	2	T45370	ribosomal protein
605	6	1.4	257	1	PNBSU2	beta-lactamase (EC	678	6	1.4	282	2	T31941	hypothetical prote
606	6	1.4	257	1	PNBS2S	beta-lactamase (EC	679	6	1.4	283	2	F86450	hypothetical prote
607	6	1.4	257	1	PKQ6	H+-transporing tw	680	6	1.4	284	2	B95306	probable regulator
608	6	1.4	257	2	C96965	transcription regu	681	6	1.4	284	2	S62839	sulfate transport
609	6	1.4	258	2	A82874	transcription anti	682	6	1.4	285	2	G75447	hypothetical prote
610	6	1.4	259	2	T51679	myb-related transc	683	6	1.4	285	2	B64045	D-arabinitol 2-deh
611	6	1.4	259	2	A70359	hydrogenase expres	684	6	1.4	285	2	E84766	probable AT-hook D
612	6	1.4	259	2	C64481	conserved hypotet	685	6	1.4	286	2	C84562	probable syntxin
613	6	1.4	259	2	E86300	protein F309.30 [i	686	6	1.4	286	2	H86664	outer membrane lip
												T31494	hypothetical prote

687	1.4	287	2	T09035	hypothetical prote	760	1.4	308	2	AH1255	conserved hypothet
688	1.4	287	2	T39689	hypothetical prote	761	1.4	308	2	T19846	hypothetical prote
689	1.4	288	2	F72164	A9R protein - vari	762	1.4	308	2	G84210	ABC transporter, A
690	1.4	288	2	T28550	hypothetical prote	763	1.4	308	2	D70875	probable PE protei
691	1.4	288	2	I36848	A8R protein - vari	764	1.4	308	2	T24732	hypothetical prote
692	1.4	288	2	T37395	probable 33.6K pro	765	1.4	308	2	A97100	probable Fe-S oxid
693	1.4	288	2	A42518	A8R protein - vacc	766	1.4	308	2	H89839	hypothetical prote
694	1.4	288	2	E83946	pyruvate synthase	767	1.4	309	2	T29293	hypothetical prote
695	1.4	289	2	AH0060	bis(5'-nucleosyl)-	768	1.4	310	1	G64079	carbamate kinase (
696	1.4	289	2	F70845	hypothetical prote	769	1.4	310	2	A84142	L-lactate dehydrog
697	1.4	289	2	H83433	hypothetical prote	770	1.4	310	2	C83483	probable 2-hydroxy
698	1.4	289	2	E97044	probable integrase	771	1.4	310	2	T26710	hypothetical prote
699	1.4	290	1	B64657	conserved hypothet	772	1.4	311	2	A56235	transcription acti
700	1.4	290	2	C71859	hypothetical prote	773	1.4	311	2	T40921	hypothetical prote
701	1.4	290	2	D81359	probable integral	774	1.4	311	2	T18705	hypothetical prote
702	1.4	290	2	T21868	hypothetical prote	775	1.4	312	2	S71207	serine O-acetyltra
703	1.4	291	2	A71194	hypothetical prote	776	1.4	312	2	A87449	conserved hypothet
704	1.4	291	2	T51067	hypothetical prote	777	1.4	313	2	C83874	arsenical pump-dri
705	1.4	291	2	A11724	ABC transporter (A	778	1.4	313	2	T33391	hypothetical prote
706	1.4	292	2	F51171	transcription fact	779	1.4	313	2	S46690	hypothetical prote
707	1.4	292	2	AH1354	ABC transporter (A	780	1.4	313	2	T39212	hypothetical prote
708	1.4	293	2	AC2215	UTP-glucose-1-phos	781	1.4	313	2	H96037	alpha-galactoside
709	1.4	293	2	D90559	gtp-binding protei	782	1.4	314	2	F96527	protein F27115.20
710	1.4	293	2	A40644	transcription regu	783	1.4	314	2	T23301	hypothetical prote
711	1.4	294	2	A57478	serine O-acetyltra	784	1.4	315	2	T06806	proline rich prote
712	1.4	294	2	D90194	hypothetical prote	785	1.4	315	2	G70165	aldose reductase h
713	1.4	294	2	T24404	hypothetical prote	786	1.4	315	2	S75143	sensory transducti
714	1.4	294	2	E69759	hypothetical prote	787	1.4	315	2	JC7572	somite Maf1 protei
715	1.4	295	2	A12683	transcription regu	788	1.4	315	2	T17340	hypothetical prote
716	1.4	295	2	B32058	mcbB protein - Esc	789	1.4	316	1	NJBW41	toxin M1-1 precurs
717	1.4	295	2	B86371	hypothetical prote	790	1.4	316	2	G97465	lysr type transcri
718	1.4	295	2	B84747	hypothetical prote	791	1.4	316	2	T34353	G protein-coupled
719	1.4	295	2	T22039	hypothetical prote	792	1.4	316	2	T08694	hypothetical prote
720	1.4	295	2	AG2858	2-dehydro-3-deoxyg	793	1.4	316	2	I64205	transcription anti
721	1.4	296	1	S55795	arginase (EC 3.5.3	794	1.4	317	2	E90811	flagellar hook-fill
722	1.4	296	2	S21306	hypothetical prote	795	1.4	317	2	A85671	flagellar hook-ass
723	1.4	297	2	E70482	Na(+) dependent tr	796	1.4	317	2	S44022	flagellar hook-ass
724	1.4	297	2	S23737	proline-rich prote	797	1.4	317	2	G70222	conserved hypothet
725	1.4	298	2	D69351	hypothetical prote	798	1.4	318	2	A99668	hypothetical prote
726	1.4	298	2	S41469	homeotic protein M	799	1.4	318	2	D85518	hypothetical prote
727	1.4	298	2	G96773	hypothetical prote	800	1.4	318	2	C64754	yaqO protein - Esc
728	1.4	299	2	T33390	hypothetical prote	801	1.4	319	1	LUBO4	annexin IV - bovin
729	1.4	299	2	T15240	hypothetical prote	802	1.4	319	2	AG0742	high-affinity zinc
730	1.4	299	2	D81384	probable lipoprote	803	1.4	319	2	S57969	repB protein - Rhi
731	1.4	300	2	T42343	hypothetical prote	804	1.4	319	2	D97768	hypothetical prote
732	1.4	300	2	T33392	hypothetical prote	805	1.4	320	2	S09208	chorion protein s3
733	1.4	300	2	G69476	hypothetical prote	806	1.4	320	2	E90597	serine/threonine-p
734	1.4	300	2	T47460	hypothetical prote	807	1.4	320	2	G83839	hypothetical prote
735	1.4	301	2	AB0164	FecCD transport fa	808	1.4	321	2	D86975	probable aspartate
736	1.4	302	2	AE2520	hypothetical prote	809	1.4	321	2	T02987	myb-related protei
737	1.4	302	2	A55641	homeotic protein G	810	1.4	321	2	B82892	conserved hypothet
738	1.4	302	2	AC2018	hypothetical prote	811	1.4	322	2	F84945	glycine-tRNA ligas
739	1.4	303	2	A56837	homeotic protein M	812	1.4	322	2	A32329	complement C3 - Af
740	1.4	303	2	B49122	homeobox protein M	813	1.4	322	2	S13955	corticolliberin-bin
741	1.4	303	2	A48130	growth arrest-spec	814	1.4	322	2	T48207	corticolliberin-bin
742	1.4	304	2	B36716	prochlorophyllid	815	1.4	322	2	S13640	transcription fact
743	1.4	304	2	A55852	membrane-associate	816	1.4	323	2	I49529	hypothetical 36K p
744	1.4	304	2	B88746	protein c18F3.3 [i	817	1.4	323	2	S56558	hypothetical 36K p
745	1.4	304	2	S59414	hypothetical prote	818	1.4	323	2	I51751	homeotic protein o
746	1.4	304	2	AD2245	hypothetical prote	819	1.4	323	2	T48160	transcription fact
747	1.4	305	2	A56554	transcription fact	820	1.4	324	2	B85064	MYB-like protein [
748	1.4	305	2	I57039	genomic screen hom	821	1.4	324	2	D86834	transcription regu
749	1.4	305	2	T24129	hypothetical prote	822	1.4	324	2	A70111	hypothetical prote
750	1.4	305	2	T25117	hypothetical prote	823	1.4	324	2	T28032	hypothetical prote
751	1.4	306	2	B64073	ribokinase (EC 2.7	824	1.4	325	2	T39169	probable oxidoredu
752	1.4	306	2	B25394	SURF-1 protein - m	825	1.4	325	2	E90168	DNA modification m
753	1.4	306	2	T09067	extensin-like prot	826	1.4	325	2	B90532	fatty acid/phospho
754	1.4	307	2	S38152	hypothetical prote	827	1.4	325	2	A37159	translation elonga
755	1.4	307	2	A45581	distal-less homeob	828	1.4	325	2	E86533	probable leader 19
756	1.4	307	2	T34143	hypothetical prote	829	1.4	325	2	E81578	conserved hypothet
757	1.4	308	2	E70392	cation efflux syst	830	1.4	325	2	A81713	conserved hypothet
758	1.4	308	2	G90057	conserved hypothet	831	1.4	326	1	S44259	sucrose operon reg
759	1.4	308	2	AE1618	conserved hypothet	832	1.4	326	2	D83483	probable metal tra

833	6	1.4	326	2	T19070	hypothetical prote	906	6	1.4	351	1	KGZQHL	histidine-rich gly
834	6	1.4	328	2	S16573	plastoquinol-plast	907	6	1.4	351	2	T32717	hypothetical prote
835	6	1.4	328	2	D72393	oligopeptide ABC t	908	6	1.4	351	2	T20270	hypothetical prote
836	6	1.4	328	2	S45998	hypothetical prote	909	6	1.4	351	2	T23423	hypothetical prote
837	6	1.4	328	2	T25231	hypothetical prote	910	6	1.4	352	2	T47820	hypothetical prote
838	6	1.4	328	2	G02469	homeotic protein D	911	6	1.4	353	2	JC7128	pleckstrin 2 - mou
839	6	1.4	329	1	S55321	cysteine synthase	912	6	1.4	353	2	H88939	protein C05E4.13 [
840	6	1.4	329	2	F91290	hypothetical prote	913	6	1.4	354	1	RGXLOA	GTP-binding regula
841	6	1.4	329	2	A86132	hypothetical prote	914	6	1.4	354	1	RGFFO2	GTP-binding regula
842	6	1.4	329	2	T45972	hypothetical prote	915	6	1.4	354	2	A61035	hypothetical prote
843	6	1.4	330	1	Q0BEP7	gene 62 protein -	916	6	1.4	354	2	A90012	homeotic protein o
844	6	1.4	330	2	C90090	ATP(GTP)-binding p	917	6	1.4	354	2	S39406	hypothetical prote
845	6	1.4	331	2	A30242	homeotic protein E	918	6	1.4	354	2	G96795	G protein alpha ch
846	6	1.4	331	2	S78452	POU-domain protein	919	6	1.4	355	2	T50479	conserved hypotet
847	6	1.4	332	1	DEHULM	L-lactate dehydrog	920	6	1.4	355	2	A70185	probable glucose-1
848	6	1.4	332	1	B39136	flagellar motor sw	921	6	1.4	355	2	A72589	otxl protein - mou
849	6	1.4	332	2	T20107	hypothetical prote	922	6	1.4	355	2	S35345	homeodomain protel
850	6	1.4	332	2	H72313	transcription regu	923	6	1.4	355	2	I58547	hypothetical prote
851	6	1.4	332	2	JH0465	homeotic protein T	924	6	1.4	355	2	C70194	hypothetical prote
852	6	1.4	333	2	A84523	probable steroid s	925	6	1.4	355	2	T15759	probable fibmbrial
853	6	1.4	334	2	A97020	isocitrate dehydro	926	6	1.4	355	2	AF0955	hypothetical prote
854	6	1.4	334	2	T22427	hypothetical prote	927	6	1.4	356	2	H72214	hypothetical prote
855	6	1.4	334	2	F97273	catabolite control	928	6	1.4	356	2	T48354	hypothetical prote
856	6	1.4	334	2	B96810	unknown protein Tl	929	6	1.4	357	2	S63146	hypothetical prote
857	6	1.4	334	2	T17213	hypothetical prote	930	6	1.4	358	2	T20286	hypothetical prote
858	6	1.4	334	2	E97635	hypothetical prote	931	6	1.4	358	2	S57176	hypothetical prote
859	6	1.4	335	2	C75119	deoxyhypusine synt	932	6	1.4	358	2	G69754	hypothetical prote
860	6	1.4	335	2	G01448	homeobox protein H	933	6	1.4	359	2	D81598	hypothetical prote
861	6	1.4	336	2	AC1979	ferric iron-bindin	934	6	1.4	359	2	C72068	hypothetical prote
862	6	1.4	336	2	B86463	AlGI-like protein,	935	6	1.4	359	2	H86555	hypothetical prote
863	6	1.4	338	2	S38030	suppressor protein	936	6	1.4	359	2	S14283	transcription fact
864	6	1.4	338	2	D98233	hypothetical prote	937	6	1.4	359	2	T21705	hypothetical prote
865	6	1.4	338	2	AG3052	DNA topoisomerase	938	6	1.4	360	2	C70773	threonine synthase
866	6	1.4	339	1	LUBO36	annexin II - bovin	939	6	1.4	360	2	D71323	conserved hypotet
867	6	1.4	339	1	LUMS36	annexin II - mouse	940	6	1.4	360	2	AE1315	histidinol-phospha
868	6	1.4	339	2	LUCH2	transcription regu	941	6	1.4	360	2	AE1687	hypothetical prote
869	6	1.4	339	2	SH2127	probable membrane	942	6	1.4	361	2	C90603	conserved hypotet
870	6	1.4	339	2	S57152	protein F10B6.6 [i	943	6	1.4	361	2	A82887	probable peptidyl-
871	6	1.4	339	2	B86281	uncharacterized pr	944	6	1.4	362	2	T38464	transcription term
872	6	1.4	339	2	E97084	hypothetical prote	945	6	1.4	362	2	G81390	probable endo-1,3-
873	6	1.4	340	2	D91016	probable protein w	946	6	1.4	363	2	G68807	hypothetical prote
874	6	1.4	340	2	T51386	VSG expression sit	947	6	1.4	364	2	T22807	hypothetical prote
875	6	1.4	340	2	A32433	D-sterospecific a	948	6	1.4	364	2	T30084	probable membrane
876	6	1.4	340	2	AD3602	glyceraldhyde-3-p	949	6	1.4	364	2	T02801	NADH2 dehydrogenas
877	6	1.4	341	1	DEKWG3	protein gp4-2 [imp	950	6	1.4	365	2	T11905	X-Pro dipeptidase
878	6	1.4	341	1	DEKWG3	metalloproteinase	951	6	1.4	365	2	JC4027	peridin-in-chloroph
879	6	1.4	341	2	A89491	probable metallopr	952	6	1.4	365	2	AD1634	conserved hypotet
880	6	1.4	341	2	T51957	hypothetical prote	953	6	1.4	365	2	S43780	probable transcrip
881	6	1.4	342	2	G84885	hypothetical prote	954	6	1.4	366	2	H87288	conserved hypotet
882	6	1.4	342	2	T18474	hypothetical prote	955	6	1.4	366	2	T34700	dihydroxyacetone k
883	6	1.4	342	2	AD2032	hypothetical prote	956	6	1.4	366	2	G70230	glucokinase (EC 2.
884	6	1.4	343	2	D89780	hypothetical prote	957	6	1.4	366	2	A99842	hypothetical prote
885	6	1.4	343	2	S52009	farnesyl-pyrophosp	958	6	1.4	366	2	G85699	probable bacteriop
886	6	1.4	343	2	A84964	biotin synthase (E	959	6	1.4	366	2	E64866	probable DNA-bindi
887	6	1.4	344	1	A40005	hoscycamine (6S)-d	960	6	1.4	367	2	D84548	transforming prote
888	6	1.4	344	2	H71825	probable histidine	961	6	1.4	367	2	AH0688	AAA-type ATPase [1
889	6	1.4	344	2	T00695	hypothetical prote	962	6	1.4	368	2	H96712	WD repeat protein
890	6	1.4	345	1	S75490	probable hydro-lysa	963	6	1.4	369	1	TVFVAF	probable ABC trans
891	6	1.4	345	2	F81751	conserved hypotet	964	6	1.4	369	2	B84356	c-Maf protein - mo
892	6	1.4	345	2	S55377	urPAB protein prec	965	6	1.4	369	2	T40905	hypothetical prote
893	6	1.4	346	2	A03308	dnak-type molecula	966	6	1.4	369	2	D81261	response regulator
894	6	1.4	346	2	S66958	UFEL protein - yea	967	6	1.4	370	2	I57555	recF protein - Str
895	6	1.4	347	2	T09140	serine O-acetyltra	968	6	1.4	370	2	T31743	probable transcrip
896	6	1.4	347	2	H91044	hypothetical prote	969	6	1.4	370	2	E97316	hypothetical prote
897	6	1.4	347	2	C85889	hypothetical prote	970	6	1.4	372	2	T09617	conserved hypotet
898	6	1.4	347	2	G97256	spore photoproduct	971	6	1.4	373	2	T10967	conserved hypotet
899	6	1.4	349	2	E95858	conserved hypotet	972	6	1.4	373	2	T52182	hypothetical prote
900	6	1.4	349	2	T31934	hypothetical prote	973	6	1.4	373	2	T20723	conserved hypotet
901	6	1.4	349	2	A96839	F23A5.2(form2) [im	974	6	1.4	374	1	A69954	hypothetical prote
902	6	1.4	349	2	B86769	GTP cyclohydrolase	975	6	1.4	374	1	G69359	conserved hypotet
903	6	1.4	349	2	AH2382	hypothetical prote	976	6	1.4	374	2	G64319	hypothetical prote
904	6	1.4	350	2	B82713	lipase modulator x	977	6	1.4	374	2	AB1943	hypothetical prote
905	6	1.4	350	2	A30046	homeotic protein r	978	6	1.4	374	2	H64500	hypothetical prote

979 6 1.4 374 2 AG3337 transcription regu  
 980 6 1.4 375 1 DECHAL alcohol dehydrogen  
 981 6 1.4 375 1 A35837 alcohol dehydrogen  
 982 6 1.4 375 2 T08572 hypothetical prote  
 983 6 1.4 375 2 C71441 probable SCARECROW  
 984 6 1.4 375 2 B82964 glycosyltransferas  
 985 6 1.4 375 2 A46390 CAMP receptor.subt  
 986 6 1.4 375 2 S49767 heat shock protein  
 987 6 1.4 376 1 C64135 ribonucleoside-dip  
 988 6 1.4 376 1 RGBY43 cell division cont  
 989 6 1.4 376 2 B72321 conserved hypothet  
 990 6 1.4 376 2 A49077 transcriptional init  
 991 6 1.4 377 2 T16763 hypothetical prote  
 992 6 1.4 377 2 AH3328 NAD(FAD)-utilizing  
 993 6 1.4 377 2 F82105 hypothetical prote  
 994 6 1.4 378 2 A49337 alanine dehydrogen  
 995 6 1.4 378 2 B90704 H repeat-associate  
 996 6 1.4 378 2 E85554 probable receptor  
 997 6 1.4 378 2 H83019 probable glycosyl  
 998 6 1.4 378 2 T21651 hypothetical prote  
 999 6 1.4 379 2 H89876 phosphoribosylamin  
 1000 6 1.4 379 2 S55245 polyubiquitin 5 -

## ALIGNMENTS

RESULT 1  
 SAZQCM  
 major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (st  
 N:Alternate names: 195K glycoprotein  
 C:Species: Plasmodium falciparum  
 C>Date: 30-Sep-1987 #sequence\_revision 31-Mar-1991 #text\_change 09-Jun-2000  
 C:Accession: A23386; S06361  
 R:Weber, J.L.; Leininger, W.M.; Lyon, J.A.  
 Nucleic Acids Res. 14, 3311-3323, 1986  
 A:Title: Variation in the gene encoding a major merozoite surface antigen of the human m  
 A:Reference number: A23386; MUID:86205236; PMID:3517809  
 A:Accession: A23386  
 A:Molecule type: DNA  
 A:Residues: 1-1104 <WEB1>  
 A:Cross-references: EMBL:X03831  
 R:Weber, J.L.; Sim, B.K.L.; Lyon, J.A.; Wolff, R.  
 Nucleic Acids Res. 16, 1206, 1988  
 A:Title: Merozoite surface protein sequence from the Camp strain of the human malaria pa  
 A:Reference number: S06361; MUID:88143999; PMID:3278296  
 A:Accession: S06361  
 A:Molecule type: DNA  
 A:Residues: 1104-1726 <WEB2>  
 A:Cross-references: EMBL:X03831  
 C:Comment: The merozoite stages of different strains have strain-specific surface antige  
 C:Comment: P. falciparum has three stages: sporozoite, merozoite, and gametocyte. The me  
 C:Superfamily: major merozoite surface antigen  
 C:Keywords: glycoprotein; merozoite; surface antigen; tandem repeat  
 F:1-19/Domain: signal sequence #status predicted <SIG>  
 F:20-1726/Product: major merozoite surface antigen #status predicted <MAT>  
 F:67-87,91-96,100-105,109-120/Region: 3-residue repeats (S-G-T)  
 F:757-765/Region: 3-residue repeats (T-E-E)  
 F:133,272,501,567,638,827,839,924,944,990,1016,1114,1221,1613,1658/Binding site: carbohy

Query Match 53.6%; Score 231; DB 1; Length 1726;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-216;  
 Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 58 AISVTMDNLGFEYDVIYKPLAGVYRSLLKQIEKNFTFNLDILNSRLKRRKY 117  
 |||||||  
 Db 1333 AISVTMDNLGFEYDVIYKPLAGVYRSLLKQIEKNFTFNLDILNSRLKRRKY 1392  
 |||||||  
 Qy 118 FLDVLESLDMQFKHSSNEYIIDSFKLNSQKNTLKSYYIKESVENDIKFAQEGIS 177  
 |||||||  
 Db 1393 FLDVLESLDMQFKHSSNEYIIDSFKLNSQKNTLKSYYIKESVENDIKFAQEGIS 1452  
 |||||||  
 Qy 178 YEKVLAKYKDDLESIKKVIKEKEKFPSPPTPPSPAKTDEQKESKFLPFLTNIETL 237

Db 1453 YEKVLAKYKDDLESIKKVIKEKEKFPSPPTPPSPAKTDEQKESKFLPFLTNIETL 1512  
 |||||||  
 Qy 238 YNNLVNKIDDYLLNLKAKINDCNVEKDEAHVKITKLSDLKIDDLFKN 288\*  
 |||||||  
 Db 1513 YNNLVNKIDDYLLNLKAKINDCNVEKDEAHVKITKLSDLKIDDLFKN 1563  
 |||||||  
 RESULT 2  
 A54498  
 major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)  
 C:Species: Plasmodium falciparum  
 C>Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 09-Jun-2000  
 C:Accession: A54498  
 R:Peterson, M.G.; Coppel, R.L.; McIntyre, P.; Langford, C.J.; Woodrow, G.; Brown, G.V  
 Mol. Biochem. Parasitol. 27, 291-302, 1988  
 A:Title: Variation in the precursor to the major merozoite surface antigens of Plasm  
 A:Reference number: A54498; MUID:88142999; PMID:2449612  
 A:Accession: A54498  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1701 <PET>  
 A:Cross-references: GB:M19143; NID:g160412; PIDN:AAA29653.1; PID:g160413  
 C:Superfamily: major merozoite surface antigen  
 C:Keywords: surface antigen

Query Match 44.1%; Score 190; DB 2; Length 1701;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-176;  
 Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 99 TFNLNLDILNSRLKRRKYFDLVLESLDMQFKHSSNEYIIDSFKLNSQKNTLLKSY 158  
 |||||||  
 Db 1349 TFNLNLDILNSRLKRRKYFDLVLESLDMQFKHSSNEYIIDSFKLNSQKNTLLKSY 1408  
 |||||||  
 Qy 159 KYIKESVENDIKFAQEGISYVEKVLAKYKDDLESIKKVIKEKEKFPSPPTPPSPAKT 218  
 |||||||  
 Db 1409 KYIKESVENDIKFAQEGISYVEKVLAKYKDDLESIKKVIKEKEKFPSPPTPPSPAKT 1468  
 |||||||  
 Qy 219 DEQKESKFLPFLTNIETLYNNLVNKIDDYLLNLKAKINDCNVEKDEAHVKITKLSDLKA 278  
 |||||||  
 Db 1469 DEQKESKFLPFLTNIETLYNNLVNKIDDYLLNLKAKINDCNVEKDEAHVKITKLSDLKA 1528  
 |||||||  
 Qy 279 IDDKIDDLFKN 288  
 |||||||  
 Db 1529 IDDKIDDLFKN 1538  
 |||||||

RESULT 3  
 A45948  
 major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)  
 C:Species: Plasmodium falciparum  
 C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 09-Jun-2000  
 C:Accession: A45948  
 R:Chang, S.P.; Kramer, K.J.; Yamaga, K.M.; Kato, A.; Case, S.E.; Siddiqui, W.A.  
 Exp. Parasitol. 67, 1-11, 1988  
 A:Title: Plasmodium falciparum: gene structure and hydropathy profile of the major me  
 A:Reference number: A45948; MUID:8900525; PMID:3049134  
 A:Accession: A45948  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1726 <CHA>  
 A:Cross-references: GB:M37213  
 C:Superfamily: major merozoite surface antigen  
 C:Keywords: surface antigen

Query Match 35.7%; Score 154; DB 2; Length 1726;  
 Best Local Similarity 100.0%; Pred. No. 3.1e-141;  
 Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 58 AISVTMDNLGFEYDVIYKPLAGVYRSLLKQIEKNFTFNLDILNSRLKRRKY 117  
 |||||||  
 Db 1333 AISVTMDNLGFEYDVIYKPLAGVYRSLLKQIEKNFTFNLDILNSRLKRRKY 1392

QY 118 FLDVLESDLMQPKHSSNEYIEDSFKLNSQKNTLLKSYKYIKESVENDIKFAQEGIS 177  
|||||  
Db 1393 FLDVLESDLMQPKHSSNEYIEDSFKLNSQKNTLLKSYKYIKESVENDIKFAQEGIS 1452  
QY 178 YYEKVLAKYKDDLESIKKVIKEEKEKFPSPPTT 211  
|||||  
Db 1453 YYEKVLAKYKDDLESIKKVIKEEKEKFPSPPTT 1486

RESULT 4  
S47282  
major merozoite surface antigen 1 - malaria parasite (Plasmodium falciparum) (strain RO-71) (fragments)  
C:Species: Plasmodium falciparum  
A:Variety: strain RO-71  
C:Date: 06-Jan-1995 #sequence\_revision 26-Jul-1996 #text\_change 09-Jun-2000  
A:Accession: S47282  
R:Tolle, R.; Bujard, H.; Cooper, J.A.  
submitted to the EMBL Data Library, July 1994  
A:Description: Plasmodium falciparum: recombination within the C-terminal region of merozoite surface antigen  
A:Reference number: S47282  
A:Accession: S47282  
A:Molecule type: DNA  
A:Residues: 1-651 <TO>  
A:Cross-references: EMBL:Z35329; NID:g535257; PIDN:CAA84558.1; PID:g535258  
A:Experimental source: strain RO-71  
C:Superfamily: major merozoite surface antigen  
C:Keywords: glycoprotein; merozoite; surface antigen

Query Match 33.6%; Score 145; DB 2; Length 651;  
Best Local Similarity 100.0%; Pred. No. 7.9e-133; Mismatches 0; Indels 0; Gaps 0;  
Matches 145; Conservative 0;

QY 58 AISTVMDNLTSGFENEYDVIYKPLAGVYRSLKQTEKNFTFNLDNLNSRLKRRKY 117  
|||||

Db 259 AISTVMDNLTSGFENEYDVIYKPLAGVYRSLKQTEKNFTFNLDNLNSRLKRRKY 318  
|||||

QY 118 FLDVLESDLMQPKHSSNEYIEDSFKLNSQKNTLLKSYKYIKESVENDIKFAQEGIS 177  
|||||

Db 319 FLDVLESDLMQPKHSSNEYIEDSFKLNSQKNTLLKSYKYIKESVENDIKFAQEGIS 378  
|||||

QY 178 YYEKVLAKYKDDLESIKKVIKEEKE 202  
|||||

Db 379 YYEKVLAKYKDDLESIKKVIKEEKE 403  
|||||

RESULT 5  
A26868  
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (strain RO-71)  
C:Species: Plasmodium falciparum  
C:Date: 19-Nov-1988 #sequence\_revision 19-Nov-1988 #text\_change 09-Jun-2000  
A:Accession: A26868  
R:Tanabe, K.; Mackay, M.; Scaife, J.G.  
J. Mol. Biol. 195, 273-287, 1987  
A:Title: Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmodium falciparum  
A:Reference number: A26868; MUID:88011243; PMID:3079521  
A:Accession: A26868  
A:Molecule type: DNA  
A:Residues: 1-1701 <TAN>  
C:Superfamily: major merozoite surface antigen  
C:Keywords: surface antigen  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-1701/Product: major merozoite surface antigen #status predicted <MAT>

Query Match 31.3%; Score 135; DB 2; Length 1701;  
Best Local Similarity 100.0%; Pred. No. 1e-122; Mismatches 135; Conservative 0; Indels 0; Gaps 0;  
Matches 135; Conservative 0;

QY 154 LKSYKYIKESVENDIKFAQEGISYVEKVLAKYKDDLESIKKVIKEEKEKFPSPPTT 213  
|||||

Db 1404 LKSYKYIKESVENDIKFAQEGISYVEKVLAKYKDDLESIKKVIKEEKEKFPSPPTT 1463  
|||||

QY 214 SPAKTEQKESKFLPFLTNIETLYNNLVNKIDYLLINLAKINCNEKDEAHVKITKL 273  
|||||

Db 1464 SPAKTEQKESKFLPFLTNIETLYNNLVNKIDYLLINLAKINCNEKDEAHVKITKL 1523  
QY 274 SDLKAIDDKIDLKFN 288  
|||||  
Db 1524 SDLKAIDDKIDLKFN 1538  
|||||

RESULT 6  
A45545  
major merozoite surface antigen - malaria parasite (Plasmodium falciparum) (fragments)  
C:Species: Plasmodium falciparum  
C:Date: 17-Feb-1994 #sequence\_revision 17-Feb-1994 #text\_change 09-Jun-2000  
A:Accession: A45545  
R:Blackman, M.J.; Ling, I.T.; Nicholls, S.C.; Holder, A.A.  
Mol. Biochem. Parasitol. 49, 29-33, 1991  
A:Title: Proteolytic processing of the Plasmodium falciparum merozoite surface protein 1  
A:Reference number: A45545; MUID:92131048; PMID:1775158  
A:Accession: A45545  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-400 <BLA>  
A:Note: Sequence extracted from NCBI backbone (NCBIN:77612, NCBIP:77621)  
C:Superfamily: major merozoite surface antigen  
C:Keywords: glycoprotein; merozoite; surface antigen; tandem repeat

Query Match 10.7%; Score 46; DB 2; Length 400;  
Best Local Similarity 100.0%; Pred. No. 1.5e-36; Mismatches 0; Indels 0; Gaps 0;  
Matches 46; Conservative 0;

QY 352 NSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCENNNGGCADA 397  
|||||

Db 301 NSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCENNNGGCADA 346  
|||||

RESULT 7  
S05603  
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)  
N:Alternate names: gp195 surface antigen  
C:Species: Plasmodium falciparum  
C:Date: 12-Feb-1993 #sequence\_revision 12-Feb-1993 #text\_change 09-Jun-2000  
A:Accession: S05603; S04850  
R:Myler, P.J.  
submitted to the EMBL Data Library, April 1989.  
A:Reference number: S05603  
A:Accession: S05603  
A:Molecule type: mRNA  
A:Residues: 1-1639 <MYL>  
A:Cross-references: EMBL:X15063; NID:g9896; PIDN:CAA33163.1; PID:g9897  
Nucleic Acids Res. 17, 5401, 1989  
A:Title: Nucleotide and deduced amino acid sequence of the gp195 (MSA-1) gene from P1  
A:Reference number: S04850; MUID:89345116; PMID:2668887  
A:Accession: S04850  
A:Molecule type: mRNA  
A:Residues: 1504-1639 <MYL2>  
A:Cross-references: EMBL:X15063  
C:Superfamily: major merozoite surface antigen  
C:Keywords: glycoprotein; merozoite; surface antigen  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-1639/Product: major merozoite surface antigen #status predicted <MAT>

Query Match 10.7%; Score 46; DB 2; Length 1639;  
Best Local Similarity 100.0%; Pred. No. 5.2e-36; Mismatches 0; Indels 0; Gaps 0;  
Matches 46; Conservative 0;

QY 352 NSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCENNNGGCADA 397  
|||||

Db 1540 NSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCENNNGGCADA 1585  
|||||

RESULT 8  
SAZQK1  
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum  
C:Date: 30-Sep-1987 #sequence\_revision 30-Sep-1987 #text\_change 09-Jun-2000  
C:Accession: A25120  
R:Mackay, M.; Goman, M.; Bone, N.; Hyde, J.E.; Scalfe, J.; Certa, U.; Stunnenberg, H.; B  
EMBO J. 4, 3823-3829, 1985  
A:Title: Polymorphism of the precursor for the major surface antigens of Plasmodium fal  
A:Reference number: A91030; MUID:86136024; PMID:3004972  
A:Accession: A25120  
A:Molecule type: DNA  
A:Residues: 1-1631 <MAC>  
C:Comment: The merozoite stages of different strains have strain-specific surface antigen  
C:Superfamily: P. falciparum has three stages: sporozoite, merozoite, and gametocyte. The m  
C:Keywords: major merozoite surface antigen  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-1631/Product: major merozoite surface antigen #status predicted <MAT>  
F:67-84/Region: 3-residue repeats (S-G-T/P)  
F:1614-1631/Domain: membrane anchor #status predicted <MBN>  
F:97,259,755,759,835,911,955,1049,1156,1163,1436,1563/Binding site: carbohydrate (Asn) (

Query Match 5.6%; Score 24; DB 1; Length 1631;  
Best Local Similarity 100.0%; Pred. No. 1.4e-14;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 NSGCFRHLDERECKCLLNKQEG 375  
|||||  
Db 1532 NSGCFRHLDERECKCLLNKQEG 1555

RESULT 9  
A24594  
probable major surface antigen (83K, 19K, 42K) precursor - malaria parasite (Plasmodium  
C:Species: Plasmodium falciparum  
C:Date: 29-Aug-1987 #sequence\_revision 29-Aug-1987 #text\_change 09-Jun-2000  
C:Accession: A24594  
R:Holder, A.A.; Lockyer, M.J.; Odink, K.G.; Sandhu, J.S.; Riveros-Moreno, V.; Nicholls,  
Nature 317, 270-273, 1985  
A:Title: Primary structure of the precursor to the three major surface antigens of Plas  
A:Reference number: A24594; MUID:86014355; PMID:2995820  
A:Accession: A24594  
A:Molecule type: DNA  
A:Residues: 1-1640 <HOL>  
C:Superfamily: major merozoite surface antigen  
C:Keywords: surface antigen

Query Match 5.3%; Score 23; DB 2; Length 1640;  
Best Local Similarity 100.0%; Pred. No. 1.3e-13;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 353 SGCFRHLDERECKCLLNKQEG 375  
|||||  
Db 1542 SGCFRHLDERECKCLLNKQEG 1564

RESULT 10  
NR80B  
pancreatic ribonuclease (EC 3.1.27.5) - American bison (tentative sequence)  
N:Alternate names: RNase 1; RNase A  
C:Species: Bison bison (American bison)  
C:Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 08-Dec-1994  
C:Accession: A91771; A90270; A94696; A00804  
R:Muskiet, F.A.J.; Welling, G.W.; Beintema, J.J.  
Int. J. Pept. Protein Res. 8, 345-348, 1976  
A:Reference number: A91771; MUID:76259396; PMID:955781  
A:Accession: A91771  
A:Molecule type: protein  
A:Residues: 1-124 <MUS>  
A:Note: comparison of peptide compositions and partial sequence determination revealed r  
R:Stewart, G.N.; Stevenson, K.J.  
Biochem. J. 135, 427-441, 1973  
A:Title: The isolation and partial characterization of ribonuclease A from Bison bison.  
A:Reference number: A90270; MUID:74081066; PMID:4772270  
A:Accession: A90270

A:Molecule type: protein  
A:Residues: 1-60,62-124 <STE>  
A:Note: peptide compositions were compared with the bovine sequence  
R:Barnard, E.A.; Cohen, M.S.; Gold, M.H.; Kim, J.K.  
Nature 240, 395-398, 1972  
A:Title: Evolution of ribonuclease in relation to polypeptide folding mechanisms.  
A:Reference number: A94696; MUID:73045781; PMID:4564316  
A:Accession: A94696  
A:Molecule type: protein  
A:Residues: 1-15,'E',17,'E',19-20,'XTA',24-30 <BAR>  
C:Superfamily: pancreatic ribonuclease  
C:Keywords: glycoprotein; hydrolase; nucleic acid digestion; pancreas  
F:12,41,119/Active site: His, Lys, His #status predicted  
F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted  
F:34/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 3.5%; Score 15; DB 1; Length 124;  
Best Local Similarity 100.0%; Pred. No. 9.2e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KETAAAKFERQHMD 33  
|||||  
Db 1 KETAAAKFERQHMD 15

RESULT 11  
NRANE  
pancreatic ribonuclease (EC 3.1.27.5) - eland  
N:Alternate names: RNase 1; RNase A  
C:Species: Tragelaphus oryx, Taurotragus oryx (eland)  
C:Date: 17-Mar-1987 #sequence\_revision 17-Mar-1987 #text\_change 31-Dec-1993  
C:Accession: A00806  
R:Russchen, F.; de Vriese, G.; Gaastra, W.; Beintema, J.J.  
Biochim. Biophys. Acta 427, 719-726, 1976  
A:Title: Studies on the covalent structure of eland pancreatic ribonuclease.  
A:Reference number: A00806; MUID:76184786; PMID:1268225  
A:Accession: A00806  
A:Molecule type: protein  
A:Residues: 1-124 <RUS>  
C:Superfamily: pancreatic ribonuclease  
C:Keywords: hydrolase; nucleic acid digestion; pancreas  
F:12,41,119/Active site: His, Lys, His #status predicted  
F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted

Query Match 3.5%; Score 15; DB 1; Length 124;  
Best Local Similarity 100.0%; Pred. No. 9.2e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KETAAAKFERQHMD 33  
|||||  
Db 1 KETAAAKFERQHMD 15

RESULT 12  
S08547  
pancreatic ribonuclease (EC 3.1.27.5) - nilgai  
C:Species: Boselaphus tragocamelus (nilgai)  
C:Date: 21-Nov-1993 #sequence\_revision 01-Dec-1995 #text\_change 05-Jun-1998  
C:Accession: S08547  
R:Beintema, J.J.  
Biochim. Biophys. Acta 621, 89-103, 1980  
A:Title: Primary structures of pancreatic ribonucleases from Bovidae. Impala, Thomson  
A:Reference number: S07141; MUID:80109825; PMID:7353035  
A:Accession: S08547  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-124 <BEI>  
C:Superfamily: pancreatic ribonuclease  
C:Keywords: hydrolase; nucleic acid digestion; pancreas  
F:12,41,119/Active site: His, Lys, His #status predicted  
F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted

Query Match 3.5%; Score 15; DB 2; Length 124;

Best Local Similarity 100.0%; Pred. No. 9.2e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KETAAKFERQHMS 33  
|||||  
Db 1 KETAAKFERQHMS 15

RESULT 13  
JC5560  
pancreatic ribonuclease (EC 3.1.27.5) A - Aspergillus niger var. macrosporus  
C:Species: Aspergillus niger var. macrosporus  
C:Date: 02-Sep-1997 #sequence\_revision 05-Sep-1997 #text\_change 07-May-1999  
C:Accession: JC5560  
R:Takahashi, K.  
Biosci. Biotechnol. Biochem. 61, 381-383, 1997  
A:Title: The specificity of peptide bond cleavage of acid proteinase A from Aspergillus  
A:Reference number: JC5560; MUID:97212043; PMID:9058983  
A:Accession: JC5560  
A:Molecule type: protein  
A:Residues: 1-124 <TAK>  
C:Superfamily: pancreatic ribonuclease  
C:Keywords: hydrolase

Query Match 3.5%; Score 15; DB 2; Length 124;  
Best Local Similarity 100.0%; Pred. No. 9.2e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KETAAKFERQHMS 33  
|||||  
Db 1 KETAAKFERQHMS 15

RESULT 14  
NR80  
pancreatic ribonuclease (EC 3.1.27.5) precursor [validated] - bovine  
N:Alternate names: ribonuclease I; ribonuclease A  
C:Species: Bos primigenius taurus (cattle)  
C:Date: 24-Apr-1984 #sequence\_revision 05-Aug-1994 #text\_change 15-Sep-2000  
C:Accession: S00897; I45998; A32471; A00804; A92018; A61293; S53757; S05528  
R:Carlsena, A.; Confalone, E.; Palmeri, M.; Libonati, M.; Furia, A.  
Nucleic Acids Res. 16, 5491-5502, 1988  
A:Title: Structure of the bovine pancreatic ribonuclease gene: the unique intervening se  
A:Reference number: S00897; MUID:86262557; PMID:2838818  
A:Accession: S00897  
A:Molecule type: DNA  
A:Residues: 1-150 <CAR>  
A:Cross-references: ENBL:X07283; NID:g671; PIDN:CAA30263.1; PID:g672  
R:Vasanthan, N.; Filipula, D.  
Gene 76, 53-60, 1989  
A:Title: Expression of bovine pancreatic ribonuclease A coded by a synthetic gene in Bac  
A:Reference number: I45998; MUID:89306659; PMID:2501158  
A:Accession: I45998  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 'M', 27-150 <VAS>  
A:Cross-references: ENBL:X15802; NID:g93; PIDN:CAA33801.1; PID:g94  
A:Note: recombinant gene expressed in E. coli  
R:Robertson, A.D.; Purisima, E.O.; Eastman, M.A.; Scheraga, H.A.  
Biochemistry 28, 5930-5938, 1989  
A:Title: Proton NMR assignments and regular backbone structure of bovine pancreatic ribo  
A:Accession: A32471; MUID:89375325; PMID:2775743  
A:Accession: A32471  
A:Molecule type: protein  
A:Residues: 27-150 <ROB>  
R:Smyle, D.G.; Stein, W.H.; Moore, S.  
J. Biol. Chem. 238, 227-234, 1963  
A:Title: The sequence of amino acid residues in bovine pancreatic ribonuclease: revision  
A:Reference number: A92016  
A:Accession: A00804  
A:Molecule type: protein  
A:Residues: 27-150 <SMY>  
A:Note: disulfide bonds were determined

R:Plummer Jr., T.H.; Hirs, C.H.W.  
J. Biol. Chem. 239, 2530-2538, 1964  
A:Title: On the structure of bovine pancreatic ribonuclease B. Isolation of a glycope  
A:Reference number: A92018  
A:Accession: A92018  
A:Molecule type: protein  
A:Residues: 27-150 <PLU>  
R:Kumagai, H.; Yoshihara, K.; Umemoto, M.; Igarashi, K.; Hirose, S.; Ohgi, K.; Irie, J.  
Biochem. 93, 865-874, 1983  
A:Title: Studies on salivary gland ribonucleases. III. Purification and properties of  
A:Reference number: A61293; MUID:83265672; PMID:6874668  
A:Accession: A61293  
A:Molecule type: protein  
A:Residues: 27-40, 'XX', 43, 'X', 45-46 <KUM>  
R:Yang, H.J.; Tsou, C.L.  
Biochem. J. 305, 379-384, 1995  
A:Title: Inactivation during denaturation of ribonuclease A by guanidinium chloride  
A:Reference number: S53757; MUID:95134214; PMID:7832749  
A:Accession: S53757  
A:Molecule type: protein  
A:Residues: 27-53; 60-63; 69-73 <YAN>  
R:Birdsall, D.L.; McPherson, A.  
submitted to the Brookhaven Protein Data Bank, August 1992  
A:Reference number: A51385; PDB:1RTB  
A:Contents: annotation: X-ray crystallography, 2.5 angstroms, residues 27-150  
R:Birdsall, D.L.; McPherson, A.  
J. Biol. Chem. 267, 22230-22236, 1992  
A:Title: Crystal structure disposition of thymidylc acid tetramer in complex with ri  
A:Reference number: A44321; MUID:93054504; PMID:1429575  
A:Contents: annotation: X-ray crystallography, 2.5 angstroms  
R:Williams, R.L.; Greene, S.M.; McPherson, A.  
submitted to the Brookhaven Protein Data Bank, September 1987  
A:Reference number: A50330; PDB:1RBB  
A:Contents: annotation: X-ray crystallography, 2.5 angstroms, residues 27-150  
R:Wlodawer, A.  
submitted to the Brookhaven Protein Data Bank, April 1995  
A:Reference number: A50927; PDB:5RSA  
A:Contents: annotation: X-ray and neutron crystallography, 2.0 angstroms, residues 27  
R:Wlodawer, A.; Bott, R.; Sjolín, L.  
J. Biol. Chem. 257, 1325-1332, 1982  
A:Title: The refined crystal structure of ribonuclease A at 2.0 angstrom resolution.  
A:Reference number: A92350; MUID:82120062; PMID:6276380  
A:Contents: annotation: X-ray crystallography, 2.0 angstroms  
R:Howlin, B.; Moss, D.S.; Harris, G.W.; Palmer, R.A.  
submitted to the Brookhaven Protein Data Bank, October 1991  
A:Reference number: A50626; PDB:3RN3  
A:Contents: annotation: X-ray crystallography, 1.45 angstroms, residues 27-150  
R:Carlisle, C.H.; Palmer, R.A.; Mazumdar, S.K.; Gorinsky, B.A.; Yeates, D.G.R.  
J. Mol. Biol. 85, 1-18, 1974  
A:Title: The structure of ribonuclease at 2.5 angstrom resolution.  
A:Reference number: A92949; MUID:74253324; PMID:4835726  
A:Contents: annotation: X-ray crystallography, 2.5 angstroms  
R:Wyckoff, H.W.; Tsernoglou, D.; Hanson, A.W.; Knox, J.R.; Lee, B.; Richards, F.M.  
J. Biol. Chem. 245, 305-328, 1970  
A:Title: The three-dimensional structure of ribonuclease-S. Interpretation of an elec  
A:Reference number: A92059; MUID:70092235; PMID:5460889  
A:Contents: annotation: X-ray crystallography, 2.0 angstroms  
R:Santoro, J.; Gonzalez, C.; Bruix, M.; Neira, J.L.; Nieto, J.L.; Herranz, J.; Rico,  
submitted to the Brookhaven Protein Data Bank, November 1992  
A:Reference number: A52045; PDB:2AAS  
A:Contents: annotation: conformation by (1)H-NMR, residues 27-150  
R:Rico, M.; Bruix, M.; Santoro, J.; Gonzalez, C.; Neira, J.L.; Nieto, J.L.; Herranz,  
Eur. J. Biochem. 183, 623-638, 1989  
A:Title: Sequential (1)H-NMR assignment and solution structure of bovine pancreatic  
A:Reference number: S05528; MUID:89377830; PMID:2776756  
A:Contents: annotation: confirmation by (1)H-NMR  
R:Shall, S.; Barnard, E.A.  
J. Mol. Biol. 41, 237-251, 1969  
A:Title: Heavy atom-labelled derivatives of bovine pancreatic ribonuclease. I. Specif  
A:Reference number: A92946; MUID:69260123; PMID:5801478  
A:Contents: annotation: active site  
R:Heinrikson, R.L.; Stein, W.H.; Crestfield, A.M.; Moore, S.  
J. Biol. Chem. 240, 2921-2934, 1965

A:Title: The reactivities of the histidine residues at the active site of ribonuclease d  
A:Reference number: A92020  
A:Contents: annotation; active site  
C:Function:  
A:Description: ribonuclease endonucleolytically hydrolyzes RNA to produce 3'-phosphomono  
A:Pathway: ribonucleic acid digestion  
A:Note: this enzyme can also destabilize or unwind the DNA helix by complexing with sing  
rginine residues of the enzyme and the phosphate groups of the nucleotides  
C:Superfamily: pancreatic ribonuclease  
C:Keywords: glycoprotein; hydrolase; nucleic acid digestion; pancreas  
F:1-26/Domain: signal sequence #status predicted <SIG>  
F:27-150/Product: pancreatic ribonuclease #status experimental <MAT>  
F:38,67,145/Active site: His, Lys, His #status experimental  
F:52-110,66-121,84-136,91-98/Disulfide bonds: #status experimental  
F:60/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 3.5% Score 15; DB 1; Length 150;  
Best Local Similarity 100.0%; Pred. No. 1.1e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KETAAKFERQHMD 33  
|||||  
DB 27 KETAAKFERQHMD 41

RESULT 15  
NRGN  
A:Title: pancreatic ribonuclease (EC 3.1.27.5) - white-tailed deer (fragment)  
N:Alternate names: RNase 1; RNase A  
C:Species: Odocoileus virginianus (white-tailed deer)  
C:Date: 05-Jun-1987 #sequence\_revision 05-Jun-1987 #text\_change 28-Apr-1993  
C:Accession: A05004  
R:Barnard, E.A.; Cohen, M.S.; Gold, M.H.; Kim, J.K.  
A:Reference number: A94696; MUID:73045781; PMID:4564316  
A:Accession: A05004  
A:Molecule type: protein  
A:Residues: 1-30 <BAR>  
A:Note: Ser-3 was also found  
C:Superfamily: pancreatic ribonuclease  
C:Keywords: hydrolase; nucleic acid digestion; pancreas

Query Match 3.2% Score 14; DB 2; Length 30;  
Best Local Similarity 100.0%; Pred. No. 2.1e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KETAAKFERQHMD 32  
|||||  
DB 1 KETAAKFERQHMD 14

RESULT 16  
NRGN  
A:Title: pancreatic ribonuclease (EC 3.1.27.5) - brindled gnu  
N:Alternate names: RNase 1; RNase A  
C:Species: Connochaetes taurinus (brindled gnu, white-bearded gnu, blue wildebeest)  
C:Date: 17-Mar-1987 #sequence\_revision 17-Mar-1987 #text\_change 04-Oct-1996  
C:Accession: A0808  
R:Groen, G.; Welling, G.W.; Beintema, J.J.  
A:Reference number: A0808  
A:Title: The amino acid sequence of gnu pancreatic ribonuclease.  
A:Reference number: A0808; MUID:76210818; PMID:1227969  
A:Accession: A0808  
A:Molecule type: protein  
A:Residues: 1-124 <GRO>  
C:Superfamily: pancreatic ribonuclease  
C:Keywords: glycoprotein; hydrolase; nucleic acid digestion; pancreas  
F:12,41,119/Active site: His, Lys, His #status predicted  
F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted  
F:34/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 2.8% Score 12; DB 1; Length 124;  
Best Local Similarity 100.0%; Pred. No. 0.00077;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Best Local Similarity 100.0%; Pred. No. 0.00077;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 AAAKFERQHMD 33  
|||||  
DB 4 AAAKFERQHMD 15

RESULT 17  
NRSH  
A:Title: pancreatic ribonuclease (EC 3.1.27.5) - sheep  
N:Alternate names: RNase 1; RNase A  
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)  
C:Date: 17-Mar-1987 #sequence\_revision 17-Mar-1987 #text\_change 04-Oct-1996  
C:Accession: A91406; A92145; A00809  
R:Welling, G.W.; Scheffer, A.J.; Beintema, J.J.  
A:Reference number: A91406; MUID:74309062; PMID:4855010  
A:Accession: A91406  
A:Molecule type: protein  
A:Residues: 1-124 <WEL>  
R:Kobayashi, R.; Hirs, C.H.W.  
J. Biol. Chem. 248, 7833-7837, 1973  
A:Title: The amino acid sequence of ovine pancreatic ribonuclease A.  
A:Reference number: A92145; MUID:74025520; PMID:4356260  
A:Accession: A92145  
A:Molecule type: protein  
A:Residues: 1-48, 'O', 50-102, 'O', 104-124 <KOB>  
C:Superfamily: pancreatic ribonuclease  
C:Keywords: glycoprotein; hydrolase; nucleic acid digestion; pancreas  
F:12,41,119/Active site: His, Lys, His #status predicted  
F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted  
F:34/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 2.8% Score 12; DB 1; Length 124;  
Best Local Similarity 100.0%; Pred. No. 0.00077;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 AAAKFERQHMD 33  
|||||  
DB 4 AAAKFERQHMD 15

RESULT 18  
NRGT  
A:Title: pancreatic ribonuclease (EC 3.1.27.5) - goat  
N:Alternate names: RNase 1; RNase A  
C:Species: Capra aegagrus hircus (domestic goat)  
C:Date: 17-Mar-1987 #sequence\_revision 17-Mar-1987 #text\_change 04-Oct-1996  
C:Accession: A00810  
R:Welling, G.W.; Scheffer, A.J.; Beintema, J.J.  
A:Reference number: A91406; MUID:74309062; PMID:4855010  
A:Accession: A00810  
A:Molecule type: protein  
A:Residues: 1-124 <WEL>  
C:Superfamily: pancreatic ribonuclease  
C:Keywords: glycoprotein; hydrolase; nucleic acid digestion; pancreas  
F:12,41,119/Active site: His, Lys, His #status predicted  
F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted  
F:34/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 2.8% Score 12; DB 1; Length 124;  
Best Local Similarity 100.0%; Pred. No. 0.00077;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 AAAKFERQHMD 33  
|||||  
DB 4 AAAKFERQHMD 15



## RESULT 19

NRPRH  
pancreatic ribonuclease (EC 3.1.27.5) - pronghorn (tentative sequence)  
N:Alternate names: RNase 1; RNase A  
C:Species: Antilocapra americana (pronghorn)  
C:Date: 28-Feb-1981 #sequence\_revision 28-Feb-1981 #text\_change 31-Mar-2000  
C:Accession: A00813  
R:Beintema, J.J.; Gastra, W.; Munnikma, J.  
J. Mol. Evol. 13, 305-316, 1979  
A:Title: Primary structure of pronghorn pancreatic ribonuclease: close relationship between  
A:Reference number: A00813; MUID:80075014; PMID:513141  
A:Accession: A00813  
A:Molecule type: protein  
A:Residues: 1-124 <BEI>  
C:Superfamily: pancreatic ribonuclease  
C:Keywords: glycoprotein; hydrolase; nucleic acid digestion; pancreas  
F:12,41,119/Active site: His, Lys, His #status predicted  
F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted  
F:34/Blinding site: carbohydrate (Asn) (covalent) (partial) #status experimental

Query Match 2.8%; Score 12; DB 1; Length 124;  
Best Local Similarity 100.0%; Pred. No. 0.00077;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KETAAAKFERQH 30

|||||

DB 1 KETAAAKFERQH 12

## RESULT 20

S07141  
pancreatic ribonuclease (EC 3.1.27.5) - impala  
C:Species: Aepyceros melampus (impala)  
C:Date: 21-Nov-1993 #sequence\_revision 01-Dec-1995 #text\_change 14-Nov-1997  
C:Accession: S07141  
R:Beintema, J.J.  
Biochim. Biophys. Acta 621, 89-103, 1980  
A:Title: Primary structures of pancreatic ribonucleases from Bovidae. Impala, Thomson's  
A:Reference number: S07141; MUID:80109825; PMID:7353035  
A:Accession: S07141  
A>Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-124 <BEI>  
C:Superfamily: pancreatic ribonuclease  
C:Keywords: glycoprotein; hydrolase; nucleic acid digestion; pancreas  
F:12,41,119/Active site: His, Lys, His #status predicted  
F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted

Query Match 2.8%; Score 12; DB 2; Length 124;  
Best Local Similarity 100.0%; Pred. No. 0.00077;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 AAAKFERQHMS 33

|||||

DB 4 AAAKFERQHMS 15

## RESULT 21

A47498  
seminal ribonuclease (EC 3.1.27.5) - synthetic  
C:Species: synthetic  
A:Note: Bos primigenius taurus (cattle) gene engineered and expressed in Escherichia coli  
C:Date: 16-Feb-1994 #sequence\_revision 30-Nov-1995 #text\_change 30-Nov-1995  
C:Accession: A47498  
R:Kim, J.S.; Raines, R.T.  
J. Biol. Chem. 268, 17392-17396, 1993  
A:Title: Bovine seminal ribonuclease produced from a synthetic gene.  
A:Reference number: A47498; MUID:93352527; PMID:7688724  
A:Accession: A47498  
A:Molecule type: nucleic acid  
A:Residues: 1-125 <KIM>  
A:Note: sequence extracted from NCBI backbone (NCBIN:136293, NCBIIP:136294)  
A:Note: part of this sequence, including the amino end of the mature protein, was determi

C:Superfamily: pancreatic ribonuclease  
 C:Keywords: homodimer; hydrolase  
 F:1-26/Domain: signal sequence #status predicted <SIG>  
 F:27-150/Product: seminal ribonuclease #status experimental <MAT>  
 F:38,67,145/Active site: His, Lys, His #status predicted  
 F:52-110,66-121,84-136,91-99/Disulfide bonds: #status experimental  
 F:57,58/Disulfide bonds: interchain #status experimental

Query Match 2.8%; Score 12; DB 1; Length 150;  
 Best Local Similarity 100.0%; Pred. No. 0.0091;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 22 AAAKFEROHMDS 33  
 |||||  
 Db 30 AAAKFERQHMDs 41

## RESULT 23

NRDEN  
 pancreatic ribonuclease (EC 3.1.27.5) - red deer (tentative sequence)  
 N:Alternate names: RNase 1; RNase A  
 C:Species: Cervus elaphus (red deer)  
 C>Date: 24-Apr-1984 #sequence\_revision 31-Dec-1991 #text\_change 03-Feb-1994  
 C:Accession: B90613; A00812  
 R:Zwiërs, H.; Scheffer, A.J.; Beintema, J.J.  
 Eur. J. Biochem. 36, 569-574, 1973  
 A:Title: Amino-acid sequences of red-deer and roe-deer pancreatic ribonucleases.  
 A:Reference number: A91207; MUID:73253236; PMID:4738402  
 A:Contents: annotation  
 A:Note: the sequence has been revised in reference A90613  
 R:Oosterhuis, S.; Welling, G.W.; Gastra, W.; Beintema, J.J.  
 Biochim. Biophys. Acta 490, 523-529, 1977  
 A:Title: Reinvestigation of the primary structures of red deer and roe deer pancreatic  
 A:Reference number: A90613; MUID:77112628; PMID:836889  
 A:Accession: B90613  
 A:Molecule type: protein  
 A:Residues: 1-124 <OOS>  
 A:Note: the authors reinvestigated the sequences of residues 1-25

C:Superfamily: pancreatic ribonuclease  
 C:Keywords: hydrolase; nucleic acid digestion; pancreas  
 F:12,41,119/Active site: His, Lys, His #status predicted  
 F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted

Query Match 2.6%; Score 11; DB 1; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 0.0073;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 22 AAAKFERQHMD 32  
 |||||  
 Db 4 AAAKFERQHMD 14

## RESULT 24

NRDEN  
 pancreatic ribonuclease (EC 3.1.27.5) - roe deer (tentative sequence)  
 N:Alternate names: RNase 1; RNase A  
 C:Species: Capreolus capreolus (roe deer)  
 C>Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 03-Feb-1994  
 C:Accession: A90613; A00812  
 R:Zwiërs, H.; Scheffer, A.J.; Beintema, J.J.  
 Eur. J. Biochem. 36, 569-574, 1973  
 A:Title: Amino-acid sequences of red-deer and roe-deer pancreatic ribonucleases.  
 A:Reference number: A91207; MUID:73253236; PMID:4738402  
 A:Contents: annotation  
 A:Note: the sequence has been revised in reference A90613  
 R:Oosterhuis, S.; Welling, G.W.; Gastra, W.; Beintema, J.J.  
 Biochim. Biophys. Acta 490, 523-529, 1977  
 A:Title: Reinvestigation of the primary structures of red deer and roe deer pancreatic  
 A:Reference number: A90613; MUID:77112628; PMID:836889  
 A:Accession: A90613  
 A:Molecule type: protein  
 A:Residues: 1-124 <OOS>  
 A:Note: the authors reinvestigated the sequences of residues 1-25 and 99

C:Superfamily: pancreatic ribonuclease  
 C:Keywords: glycoprotein; hydrolase; nucleic acid digestion; pancreas  
 F:12,41,119/Active site: His, Lys, His #status predicted  
 F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted  
 F:34/Binding site: carbohydrate (Asn) (covalent) (partial) #status experimental

Query Match 2.6%; Score 11; DB 1; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 0.0073;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 22 AAAKFERQHMD 32  
 |||||  
 Db 4 AAAKFERQHMD 14

## RESULT 25

NRDEN  
 pancreatic ribonuclease (EC 3.1.27.5) - fallow deer (tentative sequence)  
 N:Alternate names: RNase 1; RNase A  
 C:Species: Dama dama (fallow deer)  
 C>Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 31-Dec-1993  
 C:Accession: A91418; A00812  
 R:Leijenaar-van den Berg, G.; Beintema, J.J.  
 FEBS Lett. 56, 101-107, 1975  
 A:Title: The amino acid sequences of reindeer, moose and fallow deer pancreatic ribon  
 A:Reference number: A91418; MUID:76003215; PMID:1157925  
 A:Accession: A91418  
 A:Molecule type: protein  
 A:Residues: 1-124 <LEI>  
 C:Superfamily: pancreatic ribonuclease  
 C:Keywords: hydrolase; nucleic acid digestion; pancreas  
 F:12,41,119/Active site: His, Lys, His #status predicted  
 F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted

Query Match 2.6%; Score 11; DB 1; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 0.0073;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 22 AAAKFERQHMD 32  
 |||||  
 Db 4 AAAKFERQHMD 14

## RESULT 26

NRDEN  
 pancreatic ribonuclease (EC 3.1.27.5) - reindeer (tentative sequence)  
 N:Alternate names: RNase 1; RNase A  
 C:Species: Rangifer tarandus (reindeer)  
 C>Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 31-Dec-1993  
 C:Accession: B91418; A00812  
 R:Leijenaar-van den Berg, G.; Beintema, J.J.  
 FEBS Lett. 56, 101-107, 1975  
 A:Title: The amino acid sequences of reindeer, moose and fallow deer pancreatic ribon  
 A:Reference number: A91418; MUID:76003215; PMID:1157925  
 A:Accession: B91418  
 A:Molecule type: protein  
 A:Residues: 1-124 <LEI>

C:Superfamily: pancreatic ribonuclease  
 C:Keywords: hydrolase; nucleic acid digestion; pancreas  
 F:12,41,119/Active site: His, Lys, His #status predicted  
 F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted

Query Match 2.6%; Score 11; DB 1; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 0.0073;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 22 AAAKFERQHMD 32  
 |||||  
 Db 4 AAAKFERQHMD 14

## RESULT 27

NRDEN

pancreatic ribonuclease (EC 3.1.27.5) - European moose (tentative sequence)  
 N:Alternate names: RNase 1; RNase A  
 C:Species: Alces alces alces (European moose, elk)  
 C:Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 31-Dec-1993  
 C:Accession: C91418; A00812  
 R:Leijenaar-van den Berg, G.; Beintema, J.J.  
 FEBS Lett. 56, 101-107, 1975

A:Title: The amino acid sequences of reindeer, moose and fallow deer pancreatic ribonuclease  
 A:Reference number: A91418; MUID:76003215; PMID:1157925

A:Accession: C91418  
 A:Molecule type: protein  
 A:Residues: 1-124 <LET>  
 C:Superfamily: pancreatic ribonuclease  
 C:Keywords: glycoprotein; hydrolase; nucleic acid digestion; pancreas  
 F:12,41,119/Active site: His, Lys, His #status predicted  
 F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted  
 F:34/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 2.6%; Score 11; DB 1; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 0.0073;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 AAAPFERQHM 32  
 DB 4 AAAPFERQHM 14  
 |||||

## RESULT 28

NRGM

pancreatic ribonuclease (EC 3.1.27.5) - giraffe (tentative sequence)

N:Alternate names: RNase 1; RNase A  
 C:Species: Giraffa camelopardalis (giraffe)  
 C:Date: 24-Apr-1984 #sequence\_revision 24-Apr-1984 #text\_change 31-Mar-2000  
 C:Accession: A94452; A91405; A00811

R:Gaastra, W.  
 Ph.D. thesis, University of Groningen, Groningen, The Netherlands, 1975

A:Reference number: A94452  
 A:Accession: A94452  
 A:Molecule type: protein  
 A:Residues: 1-124 <GAA>  
 R:Gaastra, W.; Groen, G.W.; Beintema, J.J.  
 FEBS Lett. 41, 227-232, 1974  
 A:Title: The primary structure of giraffe pancreatic ribonuclease.  
 A:Reference number: A91405; MUID:74309061; PMID:4855009

A:Accession: A91405  
 A:Molecule type: protein  
 A:Residues: 1-124 <GA2>  
 C:Superfamily: pancreatic ribonuclease  
 C:Keywords: glycoprotein; hydrolase; nucleic acid digestion; pancreas  
 F:12,41,119/Active site: His, Lys, His #status predicted  
 F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted  
 F:34/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 2.1%; Score 9; DB 1; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 0.65;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 AAAPFERQH 30  
 DB 4 AAAPFERQH 12  
 |||||

## RESULT 29

NRGM

pancreatic ribonuclease (EC 3.1.27.5) - Arabian camel (tentative sequence)

N:Alternate names: RNase 1; RNase A  
 C:Species: Camelus dromedarius (Arabian camel)  
 C:Date: 24-Apr-1984 #sequence\_revision 17-Mar-1987 #text\_change 31-Dec-1993  
 C:Accession: A00815; A90283  
 R:Beintema, J.J.  
 FEBS Lett. 185, 115-120, 1985

A:Title: Mammalian ribonucleases. The absence of a glycosylated Asn-Pro-Thr sequence in  
 A:Reference number: A91340; MUID:85204378; PMID:3922790

A:Accession: A00815  
 A:Molecule type: protein  
 A:Residues: 1-124 <BEI>  
 R:Welling, G.W.; Groen, G.; Beintema, J.J.  
 Biochem. J. 147, 505-511, 1975  
 A:Title: The amino acid sequence of dromedary pancreatic ribonuclease.  
 A:Reference number: A90283; MUID:76039472; PMID:1167157

A:Accession: A90283  
 A:Molecule type: protein  
 A:Residues: 1-36, 'N', 38, 'X', 40-77, 'S', 79-102, 'Q', 104-124 <WEL>  
 A:Note: this sequence has been revised in reference A91340  
 C:Superfamily: pancreatic ribonuclease  
 C:Keywords: hydrolase; nucleic acid digestion; pancreas  
 F:12,41,119/Active site: His, Lys, His #status predicted  
 F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted

Query Match 2.1%; Score 9; DB 1; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 0.65;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KFERQHMS 33  
 DB 7 KFERQHMS 15  
 |||||

## RESULT 30

NRGM

pancreatic ribonuclease (EC 3.1.27.5), minor form - Arabian camel (tentative sequence)

N:Alternate names: RNase 1; RNase A  
 C:Species: Camelus dromedarius (Arabian camel)  
 C:Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 31-Dec-1993  
 C:Accession: A90229; A00815

R:Welling, G.W.; Mulder, H.; Beintema, J.J.  
 Biochem. Genet. 14, 309-317, 1976

A:Title: Allelic polymorphism in Arabian camel ribonuclease and the amino acid sequence  
 A:Reference number: A90229; MUID:76277807; PMID:962846

A:Accession: A90229  
 A:Molecule type: protein  
 A:Residues: 1-124 <WEL>  
 C:Superfamily: pancreatic ribonuclease  
 C:Keywords: hydrolase; nucleic acid digestion; pancreas  
 F:12,41,119/Active site: His, Lys, His #status predicted  
 F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted

Query Match 2.1%; Score 9; DB 1; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 0.65;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KFERQHMS 33  
 DB 7 KFERQHMS 15  
 |||||

## RESULT 31

NRGM

pancreatic ribonuclease (EC 3.1.27.5) - Bactrian camel (tentative sequence)

N:Alternate names: RNase 1; RNase A  
 C:Species: Camelus bactrianus (Bactrian camel)  
 C:Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 31-Dec-1993  
 C:Accession: B90229; A00815

R:Welling, G.W.; Mulder, H.; Beintema, J.J.  
 Biochem. Genet. 14, 309-317, 1976

A:Title: Allelic polymorphism in Arabian camel ribonuclease and the amino acid sequence  
 A:Reference number: A90229; MUID:76277807; PMID:962846

A:Accession: B90229  
 A:Molecule type: protein  
 A:Residues: 1-124 <WEL>  
 C:Superfamily: pancreatic ribonuclease  
 C:Keywords: hydrolase; nucleic acid digestion; pancreas  
 F:12,41,119/Active site: His, Lys, His #status predicted  
 F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted

Query Match 2.1%; Score 9; DB 1; Length 124;

Best Local Similarity 100.0%; Pred. No. 0.65; Mismatches 0; Indels 0; Gaps 0;  
Matches 9; Conservative 0;

QY 25 KFERQHMD 33  
DB 7 KFERQHMD 15

## RESULT 32

NRQ

pancreatic ribonuclease (EC 3.1.27.5) - muskrat (tentative sequence)

N:Alternate names: RNase 1; RNase A

C:Species: Ondatra zibethicus (muskrat)

C:Date: 24-Apr-1984 #sequence\_revision 24-Apr-1984 #text\_change 31-Mar-2000

C:Accession: A00828

R:van Dijk, H.; Sloots, B.; van den Berg, A.; Gaastra, W.; Beintema, J.J.

I:nt. J. Pept. Protein Res. 8, 305-316, 1976

A:Title: The primary structure of muskrat pancreatic ribonuclease.

A:Reference number: A00828; MUID:76212474; PMID:1279085

A:Accession: A00828

A:Molecule type: protein

A:Residues: 1-124 <VAN>

A:Note: The placement of residues 32-33 and 75-76 is tentative

C:Superfamily: pancreatic ribonuclease

C:Keywords: glycoprotein; hydrolase; nucleic acid digestion; pancreas

F:12,41,119/Active site: His, Lys, His #status predicted

F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted

F:62/Binding site: carbohydrate (Asn) (covalent) #status absent

Query Match 2.1%; Score 9; DB 1; Length 124;  
Best Local Similarity 100.0%; Pred. No. 0.65;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 25 KFERQHMD 33  
DB 7 KFERQHMD 15

## RESULT 33

NRHY

pancreatic ribonuclease (EC 3.1.27.5) - golden hamster (tentative sequence)

N:Alternate names: RNase 1; RNase A

C:Species: Mesocricetus auratus (golden hamster)

C:Date: 30-Nov-1980 #sequence\_revision 30-Nov-1980 #text\_change 31-Mar-2000

C:Accession: A00829

R:Jekel, P.A.; Sips, H.J.; Lenstra, J.A.; Beintema, J.J.

Biochimie 61, 827-839, 1979

A:Title: The amino acid sequence of hamster pancreatic ribonuclease.

A:Reference number: A00829; MUID:80088445; PMID:518928

A:Accession: A00829

A:Molecule type: protein

A:Residues: 1-124 <JEK>

C:Superfamily: pancreatic ribonuclease

C:Keywords: glycoprotein; hydrolase; nucleic acid digestion; pancreas

F:12,41,119/Active site: His, Lys, His #status predicted

F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted

F:34/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 2.1%; Score 9; DB 1; Length 124;  
Best Local Similarity 100.0%; Pred. No. 0.65;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 25 KFERQHMD 33  
DB 7 KFERQHMD 15

## RESULT 34

NRHO

pancreatic ribonuclease (EC 3.1.27.5) - horse

N:Alternate names: RNase 1; RNase A

C:Species: Equus caballus (domestic horse)

C:Date: 24-Apr-1984 #sequence\_revision 17-Mar-1987 #text\_change 31-Dec-1993

C:Accession: A91340; A00819  
R:Beintema, J.J.  
FEBS Lett. 185, 115-120, 1985

A:Title: Mammalian ribonucleases. The absence of a glycosylated Asn-Pro-Thr sequence

A:Reference number: A91340; MUID:85204378; PMID:3922790

A:Accession: A91340

A:Molecule type: protein

A:Residues: 1-128 <BEI>

R:Scheffer, A.J.; Beintema, J.J.

Eur. J. Biochem. 46, 221-233, 1974

A:Title: Horse pancreatic ribonuclease.

A:Reference number: A91222; MUID:74302367; PMID:4852291

A:Contents: annotation; tentative sequence

A:Note: this sequence has been revised

C:Superfamily: pancreatic ribonuclease

C:Keywords: glycoprotein; hydrolase; nucleic acid digestion; pancreas

F:12,41,119/Active site: His, Lys, His #status predicted

F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted

F:34/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 2.1%; Score 9; DB 1; Length 128;  
Best Local Similarity 100.0%; Pred. No. 0.66;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 25 KFERQHMD 33  
DB 7 KFERQHMD 15

## RESULT 35

NRQU

pancreatic ribonuclease (EC 3.1.27.5) - nutria (tentative sequence)

N:Alternate names: RNase 1; RNase A

C:Species: Myocastor coypus (nutria, coypu)

C:Date: 24-Apr-1984 #sequence\_revision 30-Sep-1988 #text\_change 31-Mar-2000

C:Accession: A00822

R:van den Berg, A.; van den Hende-Timmer, L.; Beintema, J.J.

Biochim. Biophys. Acta 453, 400-409, 1976

A:Title: Isolation, properties and primary structure of coypu and chinchilla pancreat

A:Reference number: A90612; MUID:77065676; PMID:999896

A:Accession: A00822

A:Molecule type: protein

A:Residues: 1-128 <VAN>

C:Superfamily: pancreatic ribonuclease

C:Keywords: glycoprotein; hydrolase; nucleic acid digestion; pancreas

F:12,41,119/Active site: His, Lys, His #status predicted

F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted

F:34/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 2.1%; Score 9; DB 1; Length 128;  
Best Local Similarity 100.0%; Pred. No. 0.66;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 25 KFERQHMD 33  
DB 7 KFERQHMD 15

## RESULT 36

NRQP

pancreatic ribonuclease (EC 3.1.27.5) - crested porcupine

C:Species: Hystrix cristata (crested porcupine)

C:Date: 17-Mar-1987 #sequence\_revision 17-Mar-1987 #text\_change 30-Sep-1993

C:Accession: A00823

R:Beintema, J.J.; Knol, G.; Martena, B.

Biochim. Biophys. Acta 705, 102-110, 1982

A:Title: The primary structures of pancreatic ribonucleases from African porcupine an

A:Reference number: A90644; MUID:83000399; PMID:7115727

A:Accession: A00823

A:Molecule type: protein

A:Residues: 1-128 <BEI>

A:Note: residues 67-78 were positioned primarily by homology with other ribonucleases

A:Note: 98-Arg.was also found

C:Superfamily: pancreatic ribonuclease  
 C:Keywords: glycoprotein; hydrolase; nucleic acid digestion; pancreas  
 F:12.41.119/Active site: His, Lys, His #status predicted  
 F:26-84.40-95-110-65-72/Disulfide bonds: #status predicted  
 F:34/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 2.1%; Score 9; DB 1; Length 128;  
 Best Local Similarity 100.0%; Pred. No. 0.66;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 25 KFERQHMD 33  
 |||||||  
 Db 7 KFERQHMD 15

## RESULT 37

S22808  
 pancreatic ribonuclease (EC 3.1.27.5) - long-tailed hamster (fragment)

C:Species: Cricetulus longicaudatus (long-tailed hamster)  
 C:Date: 22-Nov-1993 #sequence\_revision 09-Aug-1997 #text\_change 22-Jun-1999  
 C:Accession: S22808  
 R:Haugg, M.; Schein, C.H.  
 Nucleic Acids Res. 20, 612, 1992  
 A:Title: The DNA sequences of the human and hamster secretory ribonucleases determined  
 A:Reference number: S22808; MUID:92158677; PMID:1741299  
 A:Accession: S22808  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-17, 'X' 19-130 <HAU>

A:Cross-references: EMBL:X62945; NID:649519; PIDN:CAA44717.1; PID:9938047  
 A:Note: we have shown the nucleotide sequence was submitted to the EMBL Data Library, October 1991  
 C:Superfamily: pancreatic ribonuclease  
 C:Keywords: glycoprotein; hydrolase  
 F:18.47.125/Active site: His, Lys, His #status predicted  
 F:32-90.46-101.64-116.71-78/Disulfide bonds: #status predicted

Query Match 2.1%; Score 9; DB 2; Length 130;  
 Best Local Similarity 100.0%; Pred. No. 0.67;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 25 KFERQHMD 33  
 |||||||  
 Db 13 KFERQHMD 21

## RESULT 38

RSBPL5

Integrase - Staphylococcus phage L54a

C:Species: Staphylococcus phage L54a  
 A:Note: host Staphylococcus aureus  
 C:Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 23-Jul-1999  
 C:Accession: A33855  
 R:Ye, Z.H.; Lee, C.Y.

J. Bacteriol. 171, 4146-4153, 1989

A:Title: Nucleotide sequence and genetic characterization of staphylococcal bacteriophage  
 A:Reference number: A33855; MUID:89327124; PMID:2526804

A:Accession: A33855

A:Molecule type: DNA

A:Residues: 1-59 <YE2>

A:Cross-references: GB:M27965; NID:9215096; PIDN:AAA98160.1; PID:g455285

C:Genetics:

A:Gene: int

C:Superfamily: phage L54a integrase

C:Keywords: DNA binding; DNA integration

Query Match 1.9%; Score 8; DB 1; Length 59;  
 Best Local Similarity 100.0%; Pred. No. 3.2;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 160 YIKESVEN 167  
 |||||||  
 Db 12 YIKESVEN 19

## RESULT 39

C64698

probable histidine-rich metal-binding protein - Helicobacter pylori

C:Species: Helicobacter pylori

A:Variety: strains J99, 26695

C:Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 08-Oct-1999

C:Accession: C64698; C71821

R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.

Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKee,

son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey,

Nature 388, 539-547, 1997

A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser

A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.

A:Reference number: A64520; MUID:97394467; PMID:9252185

A:Accession: C64698

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-60 <TOM>

A:Cross-references: GB:AE000643; GB:AE000511; NID:92314598; PIDN:AAD08471.1; PID:g231

A:Experimental source: strain 26695

R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Dolg, P.C.; Smith, D.

Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.

Nature 397, 176-180, 1999

A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric

A:Reference number: A71800; MUID:99120557; PMID:9923682

A:Accession: C71821

A:Molecule type: DNA

A:Residues: 1-60 <ARN>

A:Cross-references: GB:AE001555; GB:AE001439; NID:g4155929; PIDN:AAD06898.1; PID:g415

A:Experimental source: strain J99

C:Genetics:

A:Gene: HP1427; jhp1320

Query Match 1.9%; Score 8; DB 2; Length 60;

Best Local Similarity 100.0%; Pred. No. 3.3;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 HHHHHHSS 9

|||||||

Db 28 HHHHHHSS 35

## RESULT 40

NRWB

pancreatic ribonuclease (EC 3.1.27.5) - domestic water buffalo

N:Alternate names: RNase 1; RNase A

C:Species: Bubalus arnee bubalis (domestic water buffalo)

C:Date: 17-Mar-1987 #sequence\_revision 17-Mar-1987 #text\_change 01-Dec-1995

C:Accession: A00805; S08548

R:Sidik, A.; Martena, B.; Beintema, J.J.

Biochem. Genet. 17, 1151-1158, 1979

A:Title: Amino acid sequence differences in pancreatic ribonucleases from water buffa

A:Reference number: A00805; MUID:80153329; PMID:540006

A:Accession: A00805

A:Molecule type: protein

A:Residues: 1-124 <SID>

A:Experimental source: Italian river breed and Indonesian swamp breed

A:Note: the sequence from the Italian river breed is shown

A:Note: the sequence from Indonesian swamp breed differs from that shown in having 34

a polymorphic form that does

R:Beintema, J.J.

Biochim. Biophys. Acta 621, 89-103, 1980

A:Title: Primary structures of pancreatic ribonucleases from Bovidae. Impala, Thomson

A:Reference number: S07141; MUID:80109825; PMID:7353035

A:Accession: S08548

A:Status: preliminary

A:Molecule type: protein

A:Residues: 1-124 <BE1>

C:Superfamily: pancreatic ribonuclease

C:Keywords: glycoprotein; hydrolase; nucleic acid digestion; pancreas

F:12.41.119/Active site: His, Lys, His #status predicted

F;26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted  
F;34/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 1.9%; Score 8; DB 1; Length 124;  
Best Local Similarity 100.0%; Pred. No. 6.1; Indels 0; Gaps 0;  
Matches 8; Conservative 0; Mismatches 0; Caps 0;

Oy 19 KETAAAKF 26  
| | | | | | | |  
Db 1 KETAAAKF 8

Search completed: May 12, 2003, 10:28:41  
Job time : 53 secs









399	6	1.4	472	6	1.4	391	1	CAJ1_YEAST	P39101	saccharomyc
400	6	1.4	473	6	1.4	391	1	HERP1_HUMAN	P22140	saccharomyc
401	6	1.4	474	6	1.4	391	1	HERP1_HUMAN	P22140	saccharomyc
402	6	1.4	475	6	1.4	391	1	PVB3_PEA	Q15011	homo sapien
403	6	1.4	476	6	1.4	391	1	PVB3_PEA	Q43064	pisum sativ
404	6	1.4	477	6	1.4	392	1	CARS_RHINI	P43232	rhizopus ni
405	6	1.4	477	6	1.4	392	1	ODD_DROME	P23803	drosophila
406	6	1.4	478	6	1.4	393	1	THL_BACSU	P54855	bacillus su
407	6	1.4	479	6	1.4	394	1	VORA_PVRFU	Q54801	pyrococcus
408	6	1.4	480	6	1.4	398	1	AAAD_HUMAN	P22760	homo sapien
409	6	1.4	481	6	1.4	398	1	CATD_CHICK	Q05744	gallus gall
410	6	1.4	482	6	1.4	400	1	CINA_ECOLI	P77808	escherichia
411	6	1.4	483	6	1.4	400	1	YD01_METJA	Q58697	methanococc
412	6	1.4	484	6	1.4	403	1	DDC_DROLE	Q96571	drosophila
413	6	1.4	485	6	1.4	403	1	MAF_HUMAN	Q27644	homo sapien
414	6	1.4	486	6	1.4	403	1	TGT_ARCFU	Q28787	archaeoglob
415	6	1.4	487	6	1.4	404	1	SNX5_HUMAN	Q9Y5X3	homo sapien
416	6	1.4	488	6	1.4	404	1	SNX5_MOUSE	Q9D8U8	homo sapien
417	6	1.4	489	6	1.4	406	1	GLYA_MYCPN	P78011	mycoplasma
418	6	1.4	490	6	1.4	406	1	NUCM_DICDI	Q23883	dictyosteli
419	6	1.4	491	6	1.4	409	1	APG3_METTH	Q27628	methanobact
420	6	1.4	492	6	1.4	409	1	CAQC_RABIT	P31235	oryctolagus
421	6	1.4	493	6	1.4	409	1	Y291_METJA	Q51739	methanococc
422	6	1.4	494	6	1.4	410	1	BR3B_HUMAN	Q12837	homo sapien
423	6	1.4	495	6	1.4	410	1	CAQC_CANFA	P12637	canis famil
424	6	1.4	496	6	1.4	411	1	BR3B_MOUSE	Q63934	mus musculu
425	6	1.4	497	6	1.4	411	1	RAPS_HUMAN	Q13702	homo sapien
426	6	1.4	498	6	1.4	412	1	DJA2_HUMAN	Q60894	homo sapien
427	6	1.4	499	6	1.4	412	1	DJA2_MOUSE	Q9QY10	mus musculu
428	6	1.4	500	6	1.4	412	1	DJA2_RAT	Q35824	rattus norv
429	6	1.4	501	6	1.4	412	1	I5P1_CANFA	Q29467	canis famil
430	6	1.4	502	6	1.4	412	1	I5P1_HUMAN	Q14642	homo sapien
431	6	1.4	503	6	1.4	413	1	CAQC_RAT	P51868	rattus norv
432	6	1.4	504	6	1.4	413	1	CP13_RAT	P05544	rattus norv
433	6	1.4	505	6	1.4	413	1	FENR_SYNY3	Q55318	synchocyst
434	6	1.4	506	6	1.4	414	1	TY11_HUMAN	P25490	homo sapien
435	6	1.4	507	6	1.4	414	1	TY11_MOUSE	Q00899	mus musculu
436	6	1.4	508	6	1.4	415	1	CAQC_MOUSE	Q09161	mus musculu
437	6	1.4	509	6	1.4	415	1	RF12_METAC	Q8T900	methanosarc
438	6	1.4	510	6	1.4	415	1	RUN3_HUMAN	Q13761	h runt-rela
439	6	1.4	511	6	1.4	415	1	YW92_MYCTU	P96897	mycobacteri
440	6	1.4	512	6	1.4	415	1	Y207_METJA	Q60269	methanococc
441	6	1.4	513	6	1.4	418	1	CP16_RAT	P09006	rattus norv
442	6	1.4	514	6	1.4	421	1	BR3A_MOUSE	P17208	mus musculu
443	6	1.4	515	6	1.4	421	1	DM3L_MOUSE	Q9CWR8	mus musculu
444	6	1.4	516	6	1.4	421	1	SYS_MYCPU	Q98RH5	mycoplasma
445	6	1.4	517	6	1.4	422	1	CSD_BORBU	Q51111	borrella bu
446	6	1.4	518	6	1.4	422	1	OMD_BOVIN	Q77742	bos taurus
447	6	1.4	519	6	1.4	423	1	AACT_HUMAN	P01011	homo sapien
448	6	1.4	520	6	1.4	423	1	BR3A_HUMAN	Q01851	homo sapien
449	6	1.4	521	6	1.4	423	1	PNT1_YEAST	P38969	saccharomyc
450	6	1.4	522	6	1.4	423	1	YJ54_YEAST	P47130	saccharomyc
451	6	1.4	523	6	1.4	425	1	L756_CABEL	Q11184	caenorhabdi
452	6	1.4	524	6	1.4	425	1	POU1_BRARE	P31366	brachydanio
453	6	1.4	525	6	1.4	425	1	YF31_METJA	Q58926	methanococc
454	6	1.4	526	6	1.4	425	1	YNN6_YEAST	P53911	saccharomyc
455	6	1.4	527	6	1.4	428	1	FXB2_MOUSE	Q64733	mus musculu
456	6	1.4	528	6	1.4	428	1	MTBA_BACAR	P19888	bacillus an
457	6	1.4	529	6	1.4	429	1	K1CR_HUMAN	P05783	homo sapien
458	6	1.4	530	6	1.4	430	1	NOR2_RAT	Q63516	rattus norv
459	6	1.4	531	6	1.4	432	1	GBAL_CRYNE	P54853	cryptococcu
460	6	1.4	532	6	1.4	432	1	GLPB_HAEIN	P43800	haemophilus
461	6	1.4	533	6	1.4	434	1	YNN6_YEAST	P53916	saccharomyc
462	6	1.4	534	6	1.4	435	1	MALC_STRPN	Q04698	streptococc
463	6	1.4	535	6	1.4	435	1	XYLA_TETHA	O82845	tetragenoco
464	6	1.4	536	6	1.4	436	1	BRNQ_HAEIN	P71345	haemophilus
465	6	1.4	537	6	1.4	437	1	PR54_YEAST	P40327	saccharomyc
466	6	1.4	538	6	1.4	437	1	2P12_BRARE	P56224	brachydanio
467	6	1.4	539	6	1.4	438	1	FUCP_ECOLI	P11551	escherichia
468	6	1.4	540	6	1.4	439	1	PR54_DROME	P48601	drosophila
469	6	1.4	541	6	1.4	439	1	SLAP_LACHE	P38059	lactobacilli
470	6	1.4	542	6	1.4	440	1	PR54_CHICK	Q90732	gallus gall
471	6	1.4	543	6	1.4	440	1	PR54_HUMAN	Q03527	homo sapien
	6	1.4	544	6	1.4	441	1	BR11_BRARE	Q90436	brachydanio



691	1	1.4	575	1	FLA2_CAMJE	P22251 campylobact	764	1	SIRI_YEAST	P21691 saccharomyc
692	6	1.4	575	1	ILVB_LACLA	Q02137 lactococcus	765	1	LAG2_YEAST	Q92325 saccharomyc
693	6	1.4	576	1	DEAF_DROME	Q24180 drosophila	766	1	EP84_HCNVA	P17151 human cytom
694	6	1.4	578	1	AC22_STRCO	P46105 streptomyce	767	1	YG04_YEAST	P53118 saccharomyc
695	6	1.4	579	1	G160_HUMAN	Q08378 homo sapien	768	1	KLC_STRPU	Q05090 strongyloce
696	6	1.4	580	1	ATPU_YEAST	P22136 saccharomyc	769	1	HDAl_SCHPO	P56523 schizosacch
697	6	1.4	585	1	YHD9_YEAST	P38732 saccharomyc	770	1	EPG_MYCCE	P47335 mycoplasma
698	6	1.4	586	1	VATA_HALVO	Q48332 halobacteri	771	1	VGLH_HSV7J	P52353 human herpe
699	6	1.4	586	1	YS73_YEAST	P46982 saccharomyc	772	1	EPGL_SYNY3	P28371 synechocyst
700	6	1.4	587	1	SYT3_MOUSE	Q35681 mus musculu	773	1	YI95_SYNY3	P74101 synechocyst
701	6	1.4	588	1	SYT3_RAT	P40748 rattus norv	774	1	HS82_ORYSA	P33126 oryza sativ
702	6	1.4	590	1	PRIM_CHLPN	Q926w4 chlamydia p	775	1	CNG3_BOVIN	Q29441 bos taurus
703	6	1.4	590	1	SYT3_HUMAN	Q9bqg1 homo sapien	776	1	SEM2_MACMU	Q95196 macaca mula
704	6	1.4	594	1	DNBK_LYCES	O04059 lycopersico	777	1	ORCL_SCHPO	P54789 schizosacch
705	6	1.4	595	1	DNBK_MYCCE	P47547 mycoplasma	778	1	GIT2_MOUSE	Q91402 mus musculu
706	6	1.4	595	1	DNBK_MYCPN	P75344 mycoplasma	779	1	VP40_HCNVA	P16753 human cytom
707	6	1.4	596	1	HNEN_ANOGA	O02491 anopheles g	780	1	CAD2_LISMO	Q60048 listeria mo
708	6	1.4	597	1	STAR_DROME	P42519 drosophila	781	1	HELS_PYRHO	Q59025 pyrococcus
709	6	1.4	597	1	SYK_CRILLO	P37879 cricetulus	782	1	EMLI_HUMAN	Q00423 homo sapien
710	6	1.4	598	1	NR42_HUMAN	P43354 homo sapien	783	1	PRDD_HUMAN	Q9h4q3 homo sapien
711	6	1.4	598	1	NR42_MOUSE	Q06219 mus musculu	784	1	RNR_MYCPN	P75529 mycoplasma
712	6	1.4	598	1	NR42_RAT	Q07917 rattus norv	785	1	EF2_ARCFU	Q28385 archaeoglob
713	6	1.4	600	1	DHSA_PARDE	Q59661 paracoccus	786	1	MYBA_XENLA	Q05935 xenopus lae
714	6	1.4	600	1	Y018_MYCCE	P47254 mycoplasma	787	1	DCOR_LACS3	P43099 lactobacill
715	6	1.4	601	1	BS4_HUMAN	Q9y5a7 homo sapien	788	1	EF2_METTE	Q93640 methanosarc
716	6	1.4	601	1	PDML_DROME	P31368 drosophila	789	1	EF2_PYRAB	Q9v128 pyrococcus
717	6	1.4	602	1	CRK_DAUCA	P53681 daucus caro	790	1	EF2_PYRUF	P29050 pyrococcus
718	6	1.4	603	1	MTM1_MOUSE	Q922c5 mus musculu	791	1	EF2_PYRHO	Q59521 pyrococcus
719	6	1.4	604	1	DNBK_BACME	P05646 bacillus me	792	1	PURL_METJA	Q58660 methanococc
720	6	1.4	604	1	XJ13_YEAST	P47030 saccharomyc	793	1	V047_HSV62	P52549 human herpe
721	6	1.4	606	1	DNBK_BACTR	Q9kws7 bacillus th	794	1	GAG_SMRVH	P21411 squirrel mo
722	6	1.4	606	1	HMID_DROAN	P22544 drosophila	795	1	YD35_MYCPN	P75443 mycoplasma
723	6	1.4	608	1	EX5A_ECOLI	P04993 escherichia	796	1	ABRA_PLAFC	P22620 plasmodium
724	6	1.4	609	1	OPA_DROME	P39768 drosophila	797	1	Y0A4_CAPEL	Q10040 caenorhabdi
725	6	1.4	610	1	DNBK_BACSH	O69268 bacillus sp	798	1	MUTA_MOUSE	P16332 mus musculu
726	6	1.4	610	1	GIDA_CHLTR	Q84506 chlamydia t	799	1	P100_HSV7J	P52519 human herpe
727	6	1.4	611	1	ADAS_DICDI	Q96759 dictyostell	800	1	ECR_LUCCU	O18531 lucilia cup
728	6	1.4	611	1	GIDA_CHLPN	Q9z7t7 chlamydia p	801	1	HUNB_DROSE	O62538 drosophila
729	6	1.4	618	1	DNBK_CYACA	Q9tlt1 cyanidium c	802	1	HUNB_DROME	P05084 drosophila
730	6	1.4	618	1	YKR4_YEAST	P36029 saccharomyc	803	1	METE_BUCAI	P57142 buchera ap
731	6	1.4	620	1	SNFL_CANAL	P52497 candida alb	804	1	SC18_YEAST	P18759 saccharomyc
732	6	1.4	621	1	CRTI_CERNC	P48537 cercospora	805	1	GIT2_HUMAN	Q14161 homo sapien
733	6	1.4	622	1	MTLI_LACLA	P35516 lactococcus	806	1	HUNB_DROYA	O62541 drosophila
734	6	1.4	626	1	NR43_HUMAN	Q92570 homo sapien	807	1	MCMT_SCHPO	O75001 schizosacch
735	6	1.4	628	1	NR43_RAT	P51179 rattus norv	808	1	DYRA_HUMAN	Q13627 homo sapien
736	6	1.4	631	1	DXS_STRCL	Q9rbn6 streptomyce	809	1	DYRA_MOUSE	Q61214 mus musculu
737	6	1.4	631	1	RPSD_BORBU	P52323 borrelia bu	810	1	DYRA_RAT	Q63470 rattus norv
738	6	1.4	632	1	ASNB_BACSU	P54420 bacillus su	811	1	IF2C_PORPU	P51257 porphyra pu
739	6	1.4	633	1	SNFL_YEAST	P06782 saccharomyc	812	1	PAG_BACAN	P13423 bacillus an
740	6	1.4	634	1	KNOB_PLAFC	P09346 plasmodium	813	1	AMYH_SACDI	P04065 saccharomyc
741	6	1.4	634	1	MET4_YEAST	P32389 saccharomyc	814	1	CULI_SCHPO	O13790 schizosacch
742	6	1.4	634	1	SELB_MOOTH	Q46455 moorella th	815	1	HUNB_DROOR	O62537 drosophila
743	6	1.4	636	1	NAFL_HUMAN	Q15025 homo sapien	816	1	AMYI_SACDI	P29760 saccharomyc
744	6	1.4	640	1	GYRB_SPICI	P34031 spiroplasma	817	1	TOPI_THEAC	Q9hm08 thermoplas
745	6	1.4	642	1	ARE2_YEAST	P53629 saccharomyc	818	1	DNK3_SYNY3	P73098 synechocyst
746	6	1.4	643	1	HS71_DROME	P02825 drosophila	819	1	YKY0_YEAST	Q02208 saccharomyc
747	6	1.4	643	1	R060_CAEEEL	Q27274 caenorhabdi	820	1	PMIP_YEAST	P35999 saccharomyc
748	6	1.4	643	1	VP4B_VACCV	P06440 vaccinia vi	821	1	ANR5_MOUSE	Q9d217 mus musculu
749	6	1.4	644	1	VP4B_VACCC	P20643 vaccinia vi	822	1	ANR5_HUMAN	Q9nu02 homo sapien
750	6	1.4	644	1	VP4B_VARV	P33818 variola vir	823	1	YG51_YEAST	P50089 saccharomyc
751	6	1.4	644	1	YEAG_ECOLI	P77391 escherichia	824	1	AD30_HUMAN	Q9ukf2 homo sapien
752	6	1.4	648	1	Y084_HUMAN	Q14699 homo sapien	825	1	LON_CAMJE	O69300 campylobact
753	6	1.4	653	1	YTH3_CAEEEL	P54002 caenorhabdi	826	1	MUTS_BUCAI	P57504 buchera ap
754	6	1.4	655	1	ABG2_HUMAN	Q9unc0 homo sapien	827	1	Y250_HUMAN	Q92540 homo sapien
755	6	1.4	656	1	UVRB_MYCCE	P47319 mycoplasma	828	1	PRIA_BACSU	P94461 bacillus su
756	6	1.4	657	1	KNOB_PLAFC	P06719 plasmodium	829	1	YL25_CAEEEL	P34418 caenorhabdi
757	6	1.4	665	1	LAMA_XENLA	P11048 xenopus lae	830	1	MUTS_THEAQ	Q56215 thermus aqu
758	6	1.4	667	1	Y366_MYCCE	P47606 mycoplasma	831	1	HUNB_DROVI	P13361 drosophila
759	6	1.4	669	1	AMY_ALTHA	P29957 alteromonas	816	1	HUNB_DROVI	P13361 drosophila
760	6	1.4	671	1	Z282_HUMAN	Q9udv7 homo sapien	817	1	MUTS_THECA	Q01778 musca domes
761	6	1.4	675	1	POLG_PRSVW	P19724 papaya ring	818	1	MUTS_THETH	Q01x6 thermus cal
762	6	1.4	677	1	BS4_MOUSE	P54729 mus musculu	818	1	YX07_CAEEEL	Q56239 thermus the
763	6	1.4	678	1	GARP_PLAFC	P13816 plasmodium	819	1	AHMI_ARATH	Q9m3h5 arabidopsis



883 6 1.4 1426 1 EGFR\_DROME  
 884 6 1.4 1431 1 DAPK\_HUMAN  
 885 6 1.4 1432 1 SKI3\_YEAST  
 886 6 1.4 1447 1 GSG1\_YEAST  
 887 6 1.4 1459 1 GEA2\_YEAST  
 888 6 1.4 1464 1 NCO2\_HUMAN  
 889 6 1.4 1466 1 SPA2\_YEAST  
 890 6 1.4 1473 1 NALI\_HUMAN  
 891 6 1.4 1483 1 UFGD4\_YEAST  
 892 6 1.4 1489 1 YPD0\_YEAST  
 893 6 1.4 1504 1 DPO2\_YEAST  
 894 6 1.4 1517 1 GLSF\_SPIOL  
 895 6 1.4 1533 1 PUM1\_DROME  
 896 6 1.4 1581 1 AROI\_PNECA  
 897 6 1.4 1586 1 SN22\_HUMAN  
 898 6 1.4 1603 1 VIT5\_CAEEL  
 899 6 1.4 1616 1 APXL\_HUMAN  
 1000 6 1.4 1629 1 RRPO\_SHMV

## ALIGNMENTS

RESULT 1  
 MSPL\_PLAFPC STANDARD; PRT; 1726 AA.  
 AC P04934;  
 DT 13-AUG-1987 (Rel. 05, Last created)  
 DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Merozoite surface protein 1 precursor (Merozoite surface antigens)  
 DE (PMMSA) (P195).

GN MSP-1.  
 OS Plasmodium falciparum (isolate Camp / Malaysia).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.  
 OX NCBI\_TaxID=5835;  
 RN [1]

SEQUENCE OF 11103 FROM N.A.  
 RX MEDLINE=86205236; PubMed=3517809;  
 Weber J.L., Leininger W.M., Lyon J.A.;  
 RT "Variation in the gene encoding a major merozoite surface antigen of  
 the human malaria parasite Plasmodium falciparum";  
 RL Nucleic Acids Res. 14:3311-3323(1986).  
 RN [2]

SEQUENCE OF 1104-1726 FROM N.A.  
 RX MEDLINE=88143999; PubMed=3278296;  
 Weber J.L., Sim B.K.L., Lyon J.A., Wolff R.;  
 RT "Merozoite surface protein sequence from the Camp strain of the human  
 malaria parasite Plasmodium falciparum";  
 RL Nucleic Acids Res. 16:1206-1206(1988).  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor  
 (Potential).

-1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42  
 kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF  
 MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; X03831; CAA27446.1;  
 CC PTR; A23386; SAZQGM.  
 DR InterPro; IPR000561; EGF-like.  
 DR Pfam; PF00008; EGF; 1.  
 KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;  
 KW Transmembrane; GPI-anchor.  
 FT SIGNAL 1 19 POTENTIAL.  
 FT CHAIN 20 1726 MEROZOITE SURFACE PROTEIN 1.  
 FT CARBOHYD 133 133 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 272 272 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 501 501 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 567 567 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 638 638 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 827 827 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 924 924 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 944 944 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 990 990 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 1016 1016 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 1114 1114 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 1221 1221 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 1613 1613 N-LINKED (GLCNAC...) (POTENTIAL).  
 SQ SEQUENCE 1726 AA; 196197 MW; DB8AD45FA352BCF3 CRC64;

Query Match 53.6%; Score 231; DB 1; Length 1726;  
 Best Local Similarity 100.0%; Pred. No. 3.5e-226;  
 Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 AISVTMDNILSGFENEYDVYILKPLAGVYRSLLKQIKOIEKNIFTFNLNLDILNSRLKRRKY 117  
 |||||  
 DB 1333 AISVTMDNILSGFENEYDVYILKPLAGVYRSLLKQIKOIEKNIFTFNLNLDILNSRLKRRKY 1392  
 |||||  
 QY 118 FLDVLESDLMQFKHISSEYIIESEKLLNSEQNTLLKSYKIKESVENDIKFAQEGIS 177  
 |||||  
 DB 1393 FLDVLESDLMQFKHISSEYIIESEKLLNSEQNTLLKSYKIKESVENDIKFAQEGIS 1452  
 |||||  
 QY 178 YERVLAKYKDDLESIRKVIKEKEKFPSPPTTPPSAKTDEQKESKFLPFLTNIETL 237  
 |||||  
 DB 1453 YERVLAKYKDDLESIRKVIKEKEKFPSPPTTPPSAKTDEQKESKFLPFLTNIETL 1512  
 |||||  
 QY 238 YNNLVNKIDYILNLIKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKN 288  
 |||||  
 DB 1513 YNNLVNKIDYILNLIKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKN 1563

## RESULT 2

MSPL\_PLAFPP STANDARD; PRT; 1726 AA.  
 AC P50495;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Merozoite surface protein 1 precursor (Merozoite surface antigens)  
 DE (PMMSA) (GP195).  
 GN MSP-1.  
 OS Plasmodium falciparum (isolate Palo Alto / Uganda).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.  
 OX NCBI\_TaxID=57270;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89005525; PubMed=3049134;  
 Chang S.P., Kramer K.J., Yamaga K.M., Kato A., Case S.E.,  
 Siddiqui W.A.;  
 RT "Plasmodium falciparum: gene structure and hydrophobic profile of the  
 major merozoite surface antigen (gp195) of the Uganda-Palo Alto  
 isolate";  
 RL Exp. Parasitol. 67:1-11(1988).  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor  
 (Potential).  
 CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42  
 kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF  
 MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.  
 CC -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).



```

RX MEDLINE-86136024; PubMed-3004972;
RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,
RA Stunnenberg H., Bujard H.;
RT "Polymorphism of the precursor for the major surface antigens of
RT Plasmodium falciparum merozoites: studies at the genetic level.";
RL EMBO J. 4:3823-3829(1985).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X05624; CAA29112.1; -
CC DR A26868; A26868.
CC DR PIR; B25120; B25120.
CC DR InterPro; IPR000561; EGF-like.
CC DR Pfam; PF00008; EGF; 1.
CC Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
CC Transmembrane; GPI-anchor.
CC -----
CC SIGNAL 1 19 POTENTIAL.
CC CHAIN 20 1701 MEROZOITE SURFACE PROTEIN 1.
CC FT CARBOHYD 110 110 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 239 239 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 470 470 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 536 536 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 607 607 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 802 802 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 899 899 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 919 919 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 965 965 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 991 991 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 1089 1089 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 1196 1196 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 1588 1588 N-LINKED (GLCNAC. .) (POTENTIAL).
CC SQ SEQUENCE 1701 AA; 193768 MW; 3FC2EC59AF96EA98 CRC64;

Query Match 44.18; Score 190; DB 1; Length 1701;
Best Local Similarity 100.0%; Pred. No. 1.6e-184;
Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 99 TFNLNLDILNSRLKRRKRYFDVLESQKPKHSSNEYIIEDESKLLNSQKNTLLKSY 158
Db 1349 TFNLNLDILNSRLKRRKRYFDVLESQKPKHSSNEYIIEDESKLLNSQKNTLLKSY 1408

Qy 159 KYIKESVENDIKFAQEGISYYEKVLAKYKDDLESIKKVIKEKEKFPSPPTTPPSPAKT 218
Db 1409 KYIKESVENDIKFAQEGISYYEKVLAKYKDDLESIKKVIKEKEKFPSPPTTPPSPAKT 1468

Qy 219 DEQKESKFLPFLNFIETLYNNLVNKIDYLYNLKAKINDCNVKEDEAHVKITKLSDLKA 278
Db 1469 DEQKESKFLPFLNFIETLYNNLVNKIDYLYNLKAKINDCNVKEDEAHVKITKLSDLKA 1528

Qy 279 IDDKIDLPKN 288
Db 1529 IDDKIDLPKN 1538

RESULT 5
MSPL_PLAF3
ID MSPL_PLAF3 STANDARD; PRT; 1682 AA.
AC P19598; Q25921;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)

```

```

DE (PMMSA) (P190).
GN MSP-1.
OS Plasmodium falciparum (isolate ro-33 / Ghana).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5834;
RN [1]
RP SEQUENCE OF 1-1061 FROM N.A.
RX MEDLINE-88166657; PubMed-3327688;
RA Certa U., Rotmann D., Matile H., Reber-Liske R.;
RT "A naturally occurring gene encoding the major surface antigen
RT precursor p190 of Plasmodium falciparum lacks tripeptide repeats.";
RL EMBO J. 6:4137-4142(1987).
RN [2]
RP SEQUENCE OF 1032-1682 FROM N.A.
RX MEDLINE-95354793; PubMed-7628566;
RA Tolle R., Bujard H., Cooper J.A.;
RT "Plasmodium falciparum: variations within the C-terminal region of
RT merozoite surface antigen-1.";
RT Exp. Parasitol. 81:47-54(1995).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M35727; AAA29715.1; -
CC EMBL; Y00087; CAA68280.1; -
CC EMBL; Z35326; CAA84555.1; -
CC PIR; S06286; S06286.
CC InterPro; IPR000561; EGF-like.
CC Pfam; PF00008; EGF; 1.
CC Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
CC Transmembrane; GPI-anchor.
CC -----
CC SIGNAL 1 19 POTENTIAL.
CC CHAIN 20 1682 MEROZOITE SURFACE PROTEIN 1.
CC FT TRANSMEM 1666 1682 MEMBRANE ANCHOR (POTENTIAL).
CC FT CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 462 462 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 528 528 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 599 599 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 785 785 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 881 881 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 901 901 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 947 947 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 1071 1071 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 1178 1178 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 1569 1569 N-LINKED (GLCNAC. .) (POTENTIAL).
CC SQ SEQUENCE 1682 AA; 192462 MW; C8ZALIE59948CAD6 CRC64;

Query Match 39.9%; Score 172; DB 1; Length 1682;
Best Local Similarity 99.7%; Pred. No. 3.2e-166;
Matches 292; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 58 AISVTMDNILSGFENEYDVLYKPLAGVYRSLKQIKBNITFPNLNLDILNSRLKRRKY 117
Db 1290 AISVTMDNILSGFENEYDVLYKPLAGVYRSLKQIKBNITFPNLNLDILNSRLKRRKY 1349

Qy 118 FLDVLESQKPKHSSNEYIIEDESKLLNSQKNTLLKSYIKESVENDIKFAQEGIS 177
Db 1350 FLDVLESQKPKHSSNEYIIEDESKLLNSQKNTLLKSYIKESVENDIKFAQEGIS 1409

Qy 178 YIEKVLAKYKDDLESIKKVIKEKEKFPSPPTTPPSPAKTDEQKESKFLPFLNFIETL 237
Db 1410 YIEKVLAKYKDDLESIKKVIKEKEKFPSPPTTPPSPAKTDEQKESKFLPFLNFIETL 1468

```



QY 238 YNNLVNKIDDYLINKAKINDCNVKEARHVKITKLSLDKALDKIDKIDKFNKPNVDFEAIKK 297  
|||||  
Db 1469 YNNLVNKIDDYLINKAKINDCNVKEARHVKITKLSLDKALDKIDKIDKFNKPNVDFEAIKK 1528  
QY 298 LINDTFTKDMGLKLLSTGLVQNPFPNTIISKLTGKFKQDMNLNISOHCVKKQCP 350  
|||||  
Db 1529 LINDTFTKDMGLKLLSTGLVQNPFPNTIISKLTGKFKQDMNLNISOHCVKKQCP 1581

RESULT 6  
MSPL\_PLAFK STANDARD; PRT: 1630 AA.  
AC P04932;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)  
DE (PMMSA) (P190).  
GN MSP-1.  
OS Plasmodium falciparum (isolate K1 / Thailand).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5839;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86136024; PubMed=3004972;  
RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,  
Stunnenberg H., Bujard H.;  
RT "Polymorphism of the precursor for the major surface antigens of  
Plasmodium falciparum merozoites: studies at the genetic level";  
RL EMBO J. 4:3823-3829(1985).  
RN [2]  
RP REVISIONS, SEQUENCE FROM N.A.  
RA Pan W., Tolle R., Bujard H.;  
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor  
CC (Potential).  
CC -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42  
CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF  
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X03371; CAA27070.1; .  
CC PIR; A25120; SAZQK1.  
CC InterPro: IPR000561; EGF-like.  
CC Pfam: PF00008; EGF: 1.  
CC Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;  
KW Transmembrane; GPI-anchor.  
FT SIGNAL 1 19 POTENTIAL.  
FT CHAIN 20 1630 MEROZOITE SURFACE PROTEIN 1.  
FT DOMAIN 67 84 TRIPEPTIDE SG(TP) REPEAT.  
FT TRANSMEM 1614 1630 MEMBRANE ANCHOR.  
FT CARBOHYD 97 97 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 259 259 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 755 755 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 759 759 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 774 774 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 835 835 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 911 911 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 955 955 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1049 1049 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1156 1156 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1165 1165 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1436 1436 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1517 1517 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT SEQUENCE 1630 AA; 187289 MW; ADBDEC3CE0A46322 CRC64;

Query Match

Query Match 10.7%; Score 46; DB 1; Length 1630;  
Best Local Similarity 100.0%; Pred. No. 3.7e-38;  
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 NSCCFRLDERECKCLLNKQEGDKCVENPNPTCENNGGCCDADA 397  
|||||  
Db 1531 NSCCFRLDERECKCLLNKQEGDKCVENPNPTCENNGGCCDADA 1576

RESULT 7  
MSPL\_PLAFW STANDARD; PRT: 1639 AA.  
AC P04933;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 01-FEB-1996 (Rel. 33, Last annotation update)  
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)  
DE (PMMSA) (P195).  
GN MSP-1.  
OS Plasmodium falciparum (isolate Wellcome).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5848;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86014355; PubMed=2995820;  
RA Holder A.A., Lockyer M.J., Odink K.G., Sandhu J.S., Riveros-Moreno V.,  
Nicholls S.C., Hillman Y., Davey L.S., Tizard M.L.V., Schwarz R.T.,  
Freeman R.R.;  
RT "Primary structure of the precursor to the three major surface  
antigens of Plasmodium falciparum merozoites";  
RL Nature 317:270-273(1985).  
RN [2]  
RP REVISIONS.  
RA Holder A.A.;  
RL Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.  
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor  
CC (Potential).  
CC -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42  
CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF  
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X02919; CAA26676.1; .  
CC PIR; A24594; A24594.  
CC InterPro: IPR000561; EGF-like.  
CC Pfam: PF00008; EGF: 1.  
CC Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;  
KW Transmembrane; GPI-anchor.  
FT SIGNAL 1 19 POTENTIAL.  
FT CHAIN 20 1639 MEROZOITE SURFACE PROTEIN 1.  
FT CARBOHYD 116 116 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 268 268 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 764 764 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 768 768 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 783 783 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 844 844 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 920 920 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 964 964 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1058 1058 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1165 1165 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1174 1174 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1445 1445 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1526 1526 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT SEQUENCE 1639 AA; 187618 MW; 2C255B6616C87F6E CRC64;

Query Match 10.7%; Score 46; DB 1; Length 1639;

```

Best Local Similarity 100.0%; Pred. No. 3.8e-38;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 352 NSGCFRHLDERECKCLLNKQEGDKCVENPNTNENNGGCDADA 397
      |||||
Db 1540 NSGCFRHLDERECKCLLNKQEGDKCVENPNTNENNGGCDADA 1585

RESULT 8
RNP_BOSTR
ID RNP_BOSTR STANDARD; PRT; 124 AA.
AC P07849;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ribonuclease pancreatic (EC 3.1.27.5) (RNase 1) (RNase A).
GN RNASE1 OR RNS1.
OS Boselaphus tragocamelus (Nilgai).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Boselaphus.
OX NCBI_TaxID=9917;
RN [1]
RP PARTIAL SEQUENCE.
RC TISSUE=Pancreas;
RX MEDLINE=80109825; PubMed=7353035;
RA Beintema J.J.;
RT "Primary structures of pancreatic ribonucleases from Bovidae. Impala,
RT Thomson's gazelle, nilgai and water buffalo.";
RL Biochim. Biophys. Acta 621:89-102(1980).
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-
CC phosphates and 3'-phosphooligonucleotides ending in C-P or U-P
CC with 2',3'-cyclic phosphate intermediates.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC RNASE1 OR RNS1.
OS Boselaphus tragocamelus (Nilgai).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Boselaphus.
OX NCBI_TaxID=9917;
RN [1]
RP PARTIAL SEQUENCE.
RC TISSUE=Pancreas;
RX MEDLINE=80109825; PubMed=7353035;
RA Beintema J.J.;
RT "Primary structures of pancreatic ribonucleases from Bovidae. Impala,
RT Thomson's gazelle, nilgai and water buffalo.";
RL Biochim. Biophys. Acta 621:89-102(1980).
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-
CC phosphates and 3'-phosphooligonucleotides ending in C-P or U-P
CC with 2',3'-cyclic phosphate intermediates.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: PANCREAS.
CC -1- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.
DR PIR: S08547; S08547.
DR HSSP: P00656; 2RNS.
DR InterPro: IPR001427; RNaseA.
DR Pfam: PF00074; RNaseA; 1.
DR PRINTS: PR00794; RIBONUCLEASE.
DR PRODOM: PD000535; RNaseA; 1.
DR SMART: SM00092; RNase_Pc; 1.
DR PROSITE: PS00127; RNASE_PANCREATIC; 1.
KW Hydrolase; Nuclease; Endonuclease.
FT DISULFID 26 84 BY SIMILARITY.
FT DISULFID 40 95 BY SIMILARITY.
FT DISULFID 58 110 BY SIMILARITY.
FT DISULFID 65 72 BY SIMILARITY.
FT ACT_SITE 12 12 BY SIMILARITY.
FT ACT_SITE 41 41 BY SIMILARITY.
FT ACT_SITE 119 119 BY SIMILARITY.
SQ SEQUENCE 124 AA; 13685 MW; 09E84341868BC94C CRC64;

Query Match 3.5%; Score 15; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 1.1e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19 KETAAKFERQHMS 33
      |||||
Db 1 KETAAKFERQHMS 15

RESULT 9
RNP_TRAOR
ID RNP_TRAOR STANDARD; PRT; 124 AA.
AC P00658;
DT 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ribonuclease pancreatic (EC 3.1.27.5) (RNase 1) (RNase A).
GN RNASE1 OR RNS1.
OS Tragelaphus oryx (Eland) (Taurotragus oryx).

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Tragelaphus.
OX NCBI_TaxID=9945;
RN [1]
RP PARTIAL SEQUENCE.
RC TISSUE=Pancreas;
RX MEDLINE=76184786; PubMed=1268225;
RA Russchen F., de Vrieze G., Gastra W., Beintema J.J.;
RT "Studies on the covalent structure of eland pancreatic ribonuclease.";
RL Biochim. Biophys. Acta 427:719-726(1976).
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-
CC phosphates and 3'-phosphooligonucleotides ending in C-P or U-P
CC with 2',3'-cyclic phosphate intermediates.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: PANCREAS.
CC -1- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.
DR PIR: A00806; NRANE.
DR HSSP: P00656; 2RNS.
DR InterPro: IPR001427; RNaseA.
DR Pfam: PF00074; RNaseA; 1.
DR PRINTS: PR00794; RIBONUCLEASE.
DR PRODOM: PD000535; RNaseA; 1.
DR SMART: SM00092; RNase_Pc; 1.
DR PROSITE: PS00127; RNASE_PANCREATIC; 1.
KW Hydrolase; Nuclease; Endonuclease.
FT DISULFID 26 84 BY SIMILARITY.
FT DISULFID 40 95 BY SIMILARITY.
FT DISULFID 58 110 BY SIMILARITY.
FT DISULFID 65 72 BY SIMILARITY.
FT ACT_SITE 12 12 BY SIMILARITY.
FT ACT_SITE 41 41 BY SIMILARITY.
FT ACT_SITE 119 119 BY SIMILARITY.
SQ SEQUENCE 124 AA; 13742 MW; E612512E71A5D00A CRC64;

Query Match 3.5%; Score 15; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 1.1e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19 KETAAKFERQHMS 33
      |||||
Db 1 KETAAKFERQHMS 15

RESULT 10
RNP_BOVIN
ID RNP_BOVIN STANDARD; PRT; 150 AA.
AC P00656;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ribonuclease pancreatic precursor (EC 3.1.27.5) (RNase 1) (RNase A).
GN RNASE1 OR RNS1.
OS Bos taurus (Bovine), and
OS Bison bison (American bison).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913, 9901;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=Bovine;
RX MEDLINE=86262557; PubMed=2838818;
RA Carsana A., Confalone E., Palmieri M., Libonati M., Furia A.;
RT "Structure of the bovine pancreatic ribonuclease gene: the unique
RT intervening sequence in the 5' untranslated region contains a
RT promoter-like element.";
PL Nucleic Acids Res. 16:5491-5502(1988).
RN [2]
RP SEQUENCE OF 23-150 FROM N.A.
RC SPECIES=Bovine;
RX MEDLINE=9603920; PubMed=7479688;
RA Delcardayre S.B., Ribo M., Quirk D.J., Rutter W.J.,

```

RA Raines R.T.;  
RT "Engineering ribonuclease A: production, purification and  
RL characterization of wild-type enzyme and mutants at Gln11.";  
RN Protein Eng. 8:261-273(1995).  
RX [3]  
RP SEQUENCE OF 27-150, AND DISULFIDE BONDS.  
RC SPECIES=Bovine;  
RA Smyth D.G., Stein W.H., Moore S.;  
RT "The sequence of amino acid residues in bovine pancreatic  
RN ribonuclease: revisions and confirmations.";  
RL J. Biol. Chem. 238:227-234(1963).  
RX [4]  
RP SEQUENCE OF 27-150.  
RC SPECIES=Bovine;  
RA Plummer T.H., Jr., Hirs C.H.W.;  
RT "On the structure of bovine pancreatic ribonuclease B. Isolation of  
RN a glycopeptide.";  
RL J. Biol. Chem. 239:2530-2538(1964).  
RX [5]  
RP ACTIVE SITE.  
RC SPECIES=Bovine;  
RA Heinrichson R.L., Stein W.H., Crestfield A.M., Moore S.;  
RT "The reactivities of the histidine residues at the active site of  
RN ribonuclease toward halo acids of different structures.";  
RL J. Biol. Chem. 240:2921-2934(1965).  
RX [6]  
RP ACTIVE SITE.  
RC SPECIES=Bovine;  
RA Shall S., Barnard E.A.;  
RT "Heavy atom-labelled derivatives of bovine pancreatic ribonuclease.  
RN I. Specific reactions of ribonuclease with N-acetylhomocysteine  
thiolactone and silver ion.";  
RL J. Mol. Biol. 41:237-251(1969).  
RX [7]  
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
RC SPECIES=Bovine;  
RA Wyckoff H.W., Tsernoglou D., Hanson A.W., Knox J.R., Lee B.,  
RT Richards F.M.;  
RN "The three-dimensional structure of ribonuclease-S. Interpretation of  
an electron density map at a nominal resolution of 2 A.";  
RL J. Biol. Chem. 245:305-328(1970).  
RX [8]  
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).  
RC SPECIES=Bovine;  
RA Carlisle C.H., Palmer R.A., Mazumdar S.K., Gorinsky B.A.,  
RT Yeates D.G.R.;  
RN "The structure of ribonuclease at 2.5-A resolution.";  
RL J. Mol. Biol. 85:1-18(1974).  
RX [9]  
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
RC SPECIES=Bovine;  
RA Wlodawer A., Bott R., Sjoelin L.;  
RT "The refined crystal structure of ribonuclease A at 2.0-A  
RN resolution.";  
RL J. Biol. Chem. 257:1325-1332(1982).  
RX [10]  
RP X-RAY CRYSTALLOGRAPHY (1.26 ANGSTROMS).  
RC SPECIES=Bovine;  
RA Wlodawer A., Svensson L.A., Sjoelin L., Gilliland G.L.;  
RT "Structure of phosphate-free ribonuclease A refined at 1.26 A.";  
RN Biochemistry 27:2705-2717(1988).  
RX [11]  
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).  
RC SPECIES=Bovine;  
RA Leonidas D.D., Shapiro R., Irons L.I., Russo N., Acharya K.R.;  
RT "Crystal structures of ribonuclease A complexes with 5'-  
RN diphosphoadenosine 3'-phosphate and 5'-diphosphoadenosine  
2'-phosphate at 1.7-A resolution.";

RL Biochemistry 36:5578-5588(1997).  
RN [12]  
RP STRUCTURE BY NMR.  
RC SPECIES=Bovine;  
RX MEDLINE=89375325; PubMed=2775743;  
RA Robertson A.D., Purisma E.O., Eastman M.A., Scheraga H.A.;  
RT "Proton NMR assignments and regular backbone structure of bovine  
RN pancreatic ribonuclease A in aqueous solution.";  
RL Biochemistry 28:5930-5938(1989).  
RX [13]  
RP STRUCTURE BY NMR.  
RC SPECIES=Bovine;  
RX MEDLINE=89377830; PubMed=2776756;  
RA Rico M., Bruix M., Santoro J., Gonzalez C., Neira J.L., Nieto J.L.,  
RT Herranz J.;  
RN "Sequential 1H-NMR assignment and solution structure of bovine  
pancreatic ribonuclease A.";  
RL Eur. J. Biochem. 183:623-638(1989).  
RX [14]  
RP STRUCTURE BY NMR.  
RC SPECIES=Bovine;  
RX MEDLINE=93044359; PubMed=1841699;  
RA Rico M., Santoro J., Gonzalez C., Bruix M., Neira J.L., Nieto J.L.,  
RT Herranz J.;  
RN "3D structure of bovine pancreatic ribonuclease A in aqueous  
solution: an approach to tertiary structure determination from a  
small basis of 1H NMR NOE correlations.";  
RL J. Biomol. NMR 1:283-298(1991).  
RX [15]  
RP DNA-BINDING.  
RC SPECIES=Bovine;  
RX MEDLINE=86179900; PubMed=3961503;  
RA McPherson A., Brayer G., Cascio D., Williams R.;  
RT "The mechanism of binding of a polynucleotide chain to pancreatic  
RN ribonuclease.";  
RL Science 232:765-768(1986).  
RX [16]  
RP SEQUENCE OF 27-150.  
RC SPECIES=B. bison;  
RX MEDLINE=76259396; PubMed=955781;  
RA Muskiet F.A.J., Welling G.W., Beintema J.J.;  
RT "Studies on the primary structure of bison pancreatic ribonuclease.";  
RN Int. J. Pept. Protein Res. 8:345-348(1976).  
RX [17]  
RP CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-  
phosphates and 3'-phosphooligonucleotides ending in C-P or U-P  
with 2',3'-cyclic phosphate intermediates.  
RN [18]  
RP SUBCELLULAR LOCATION: Secreted.  
RX [19]  
RP TISSUE SPECIFICITY: PANCREAS.  
RN [20]  
RP MISCELLANEOUS: RIBONUCLEASE CAN DESTABILIZE OR UNWIND THE DNA  
HELIX BY COMPLEXING WITH SINGLE-STRANDED DNA; THIS COMPLEX ARISES  
BY AN EXTENDED MULTISITE CATION-ANTION INTERACTION BETWEEN THE  
LYSINE AND ARGININE RESIDUES OF THE ENZYME AND THE PHOSPHATE  
GROUPS OF THE NUCLEOTIDES.  
RN [21]  
RP MISCELLANEOUS: THE BISON SEQUENCE APPEARS TO BE IDENTICAL WITH  
THAT OF BOVINE.  
RX [22]  
RP SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.  
RN [23]  
RP DATABASE: NAME=worthington enzyme manual;  
WWW="http://www.worthington-biochem.com/manual/R/RNASE.html".  
RX [24]  
RP This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
RN [25]  
RP EMBL: X07283; CAA30263.1;  
RX EMBL: S80747; AAB35594.1;  
RN PIR: A00804; NRBO.  
RX PIR: A91771; NRBOB.  
RN PIR: A32471; A32471.  
RX PIR: S00897; S00897.

```

DR PIR: S05528; S05528.
DR PDB: 1RBB; 15-JUL-92.
DR PDB: 3RN3; 15-OCT-91.
DR PDB: 2RNS; 31-JAN-94.
DR PDB: 1RSM; 15-OCT-89.
DR PDB: 1SRN; 13-JUL-92.
DR PDB: 5RSA; 15-APR-93.
DR PDB: 6RSA; 15-OCT-89.
DR PDB: 7RSA; 15-JAN-93.
DR PDB: 8RSA; 15-APR-91.
DR PDB: 9RSA; 15-APR-91.
DR PDB: 1RAT; 13-JUL-93.
DR PDB: 2RAT; 15-JUL-93.
DR PDB: 3RAT; 15-JUL-93.
DR PDB: 4RAT; 15-JUL-93.
DR PDB: 5RAT; 15-JUL-93.
DR PDB: 6RAT; 15-JUL-93.
DR PDB: 7RAT; 15-JUL-93.
DR PDB: 8RAT; 15-JUL-93.
DR PDB: 9RAT; 15-JUL-93.
DR PDB: 1RBC; 31-OCT-93.
DR PDB: 1RBD; 31-OCT-93.
DR PDB: 1RBE; 31-OCT-93.
DR PDB: 1RBF; 31-OCT-93.
DR PDB: 1RBG; 31-OCT-93.
DR PDB: 1RBH; 31-OCT-93.
DR PDB: 1RBI; 31-OCT-93.

Query Match 3.5%; Score 15; DB 1; Length 150;
Best Local Similarity 100.0%; Pred. No. 1.3e-07; Indels 0; Gaps 0;
Matches 15; Conservative 0; Mismatches 0;

Qy 19 KETAAKFERQHMD 33
Db 27 KETAAKFERQHMD 41

RESULT 11
RNP_ODOVI
ID RNP_ODOVI STANDARD; PRT; 30 AA.
AC P19640;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ribonuclease pancreatic (EC 3.1.27.5) (RNase 1) (Fragment).
GN RNASE1 OR RNS1.
OS Odocoileus virginianus virginianus (Virginia white-tailed deer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Cervoidae;
OC Cervidae; Odocoileinae; Odocoileus.
OX NCBI_TaxID=9875;
RN [1]
RP SEQUENCE.
RC TISSUE=Pancreas;
RX MEDLINE=73045781; PubMed=4564316;
RA Bernard E.A., Cohen M.S., Gold M.H., Lim J.K.;
RT "Evolution of ribonuclease in relation to polypeptide folding mechanisms."
RL Nature 240:395-398(1972).
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-phosphates and 3'-phosphooligonucleotides ending in C-P or U-P with 2',3'-cyclic phosphate intermediates.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: PANCREAS.
CC -1- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.
DR PIR: A05004; A05004.
DR HSP; P00656; 2RNS.
DR InterPro: IPR001427; RNaseA.
DR Pfam: PF00074; rnasea; 1.
DR PROSITE: PS00127; RNASE_PANCREATIC; PARTIAL.
KW Hydrolase; Nuclease; Endonuclease.
FT ACT_SITE 12 12
FT VARIANT 3 3 T -> S.

```

```

FT NON_TER 30 30
SQ SEQUENCE 30 AA: 9184E7C816394D20 CRC64;

Query Match 3.2%; Score 14; DB 1; Length 30;
Best Local Similarity 100.0%; Pred. No. 3.1e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19 KETAAKFERQHMD 32
Db 1 KETAAKFERQHMD 14

RESULT 12
RNP_AEPME
ID RNP_AEPME STANDARD; PRT; 124 AA.
AC P07847;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ribonuclease pancreatic (EC 3.1.27.5) (RNase 1) (RNase A).
GN RNASE1 OR RNS1.
OS Aepyceros melampus (Impala).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidae;
OC Bovidae; Aepycerotinae; Aepyceros.
OX NCBI_TaxID=9897;
RN [1]
RP SEQUENCE.
RC TISSUE=Pancreas;
RX MEDLINE=80109825; PubMed=7353035;
RA Beintema J.J.;
RT "Primary structures of pancreatic ribonucleases from Bovidae. Impala, Thomson's gazelle, nilgai and water buffalo."
RL Biochim. Biophys. Acta 621:89-102(1980).
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-phosphates and 3'-phosphooligonucleotides ending in C-P or U-P with 2',3'-cyclic phosphate intermediates.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: PANCREAS.
CC -1- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.
DR PIR: S07141; S07141.
DR HSP; P00656; 1SRN.
DR InterPro: IPR001427; RNaseA.
DR Pfam: PF00074; rnasea; 1.
DR PRINTS; PR00794; RIBONUCLEASE.
DR PRODOM; PD000535; RNaseA; 1.
DR SMART; SM00092; RNase_Pc; 1.
DR PROSITE; PS00127; RNASE_PANCREATIC; 1.
KW Hydrolase; Nuclease; Endonuclease; Glycoprotein.
FT CARBOHYD 34 34 N-LINKED (GLCNAC... ) (PARTIAL).
FT DISULFID 26 84 BY SIMILARITY.
FT DISULFID 40 95 BY SIMILARITY.
FT DISULFID 58 110 BY SIMILARITY.
FT DISULFID 65 72 BY SIMILARITY.
FT ACT_SITE 12 12 BY SIMILARITY.
FT ACT_SITE 41 41 BY SIMILARITY.
FT ACT_SITE 119 119 BY SIMILARITY.
SQ SEQUENCE 124 AA: 13678 MW; B3018B023584C5F5 CRC64;

Query Match 2.8%; Score 12; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.00012;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 22 AAARFERQHMD 33
Db 4 AAARFERQHMD 15

RESULT 13
RNP_ANTAM
ID RNP_ANTAM STANDARD; PRT; 124 AA.
AC P00668;
DT 21-JUL-1986 (Rel. 01, Created)

```

DT 21-JUL-1986 (Rel. 01, Last sequence update) .  
 DE 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Ribonuclease pancreatic (EC 3.1.27.5) (RNase 1) (RNase A).  
 GN RNASE1 OR RNS1.  
 OS Antilocapra americana (Pronghorn).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Antilocapridae; Antilocapra.  
 OX NCBI\_TaxID=9891;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Pancreas;  
 RX MEDLINE=80075014; PubMed=513141;  
 RA Beintema J.J., Gastra W., Munniksma J.;  
 RT "Primary structure of pronghorn pancreatic ribonuclease: close  
 relationship between giraffe and pronghorn.";  
 RL J. Mol. Evol. 13:305-316(1979).  
 CC -|- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-  
 phosphates and 3'-phosphooligonucleotides ending in C-P or U-P  
 with 2',3'-cyclic phosphate intermediates.  
 CC -|- SUBCELLULAR LOCATION: Secreted.  
 CC -|- TISSUE SPECIFICITY: PANCREAS.  
 CC -|- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.  
 DR PIR: A00813; NRPB.  
 DR HSP: P00656; IRBG.  
 DR InterPro: IPR001427; RNaseA.  
 DR Pfam: PF00074; RNaseA; 1.  
 DR PRINTS: PR00794; RIBONUCLEASE.  
 DR PRODOM: PD000535; RNaseA; 1.  
 DR SMART: SM00092; RNase\_Pc; 1.  
 DR PROSITE: PS00127; RNASE\_PANCREATIC; 1.  
 KW Hydrolyase; Nuclease; Endonuclease; Glycoprotein.  
 FT DISULFID 26 84 BY SIMILARITY.  
 FT DISULFID 40 95 BY SIMILARITY.  
 FT DISULFID 58 110 BY SIMILARITY.  
 FT DISULFID 65 72 BY SIMILARITY.  
 FT ACT\_SITE 12 12 BY SIMILARITY.  
 FT ACT\_SITE 41 41 BY SIMILARITY.  
 FT ACT\_SITE 119 119 BY SIMILARITY.  
 FT CARBOHYD 34 34 N-LINKED (GLCNAC...) (PARTIAL).  
 SQ SEQUENCE 124 AA; 13711 MW; 9435EF532420F852 CRC64;

Query Match 2.8%; Score 12; DB 1; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 0.00012;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KETAARKFERQH 30  
 |||||  
 DB 1 KETAARKFERQH 12

RESULT 14  
 RNP\_CONTA  
 ID RNP\_CONTA STANDARD; PRT; 124 AA.  
 AC P00660;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Ribonuclease pancreatic (EC 3.1.27.5) (RNase 1) (RNase A).  
 GN RNASE1 OR RNS1.  
 OS Connochaetes taurinus (Brindled gnu).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Alcelaphinae; Connochaetes.  
 OX NCBI\_TaxID=9927;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Pancreas;  
 RX MEDLINE=76210818; PubMed=1227969;  
 RA Groen G., Wellling G.W., Beintema J.J.;  
 RT "The amino acid sequence of gnu pancreatic ribonuclease.";  
 RL FEBS Lett. 60:300-304(1975).  
 CC -|- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-

CC phosphates and 3'-phosphooligonucleotides ending in C-P or U-P  
 with 2',3'-cyclic phosphate intermediates.  
 CC -|- SUBCELLULAR LOCATION: Secreted.  
 CC -|- TISSUE SPECIFICITY: PANCREAS.  
 CC -|- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.  
 DR PIR: A00808; NRGN.  
 DR HSP: P00656; ISRN.  
 DR InterPro: IPR001427; RNaseA.  
 DR Pfam: PF00074; RNaseA; 1.  
 DR PRINTS: PR00794; RIBONUCLEASE.  
 DR PRODOM: PD000535; RNaseA; 1.  
 DR SMART: SM00092; RNase\_Pc; 1.  
 DR PROSITE: PS00127; RNASE\_PANCREATIC; 1.  
 KW Hydrolyase; Nuclease; Endonuclease.  
 FT DISULFID 26 84  
 FT DISULFID 40 95  
 FT DISULFID 58 110  
 FT DISULFID 65 72  
 FT ACT\_SITE 12 12  
 FT ACT\_SITE 41 41  
 FT ACT\_SITE 119 119  
 SQ SEQUENCE 124 AA; 13686 MW; 3564BE9D9D871195 CRC64;

Query Match 2.8%; Score 12; DB 1; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 0.00012;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 AAARKFERQHMS 33  
 |||||  
 DB 4 AAARKFERQHMS 15

RESULT 15  
 RNP\_SHEEP  
 ID RNP\_SHEEP STANDARD; PRT; 124 AA.  
 AC P00661; P04420;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Ribonuclease pancreatic (EC 3.1.27.5) (RNase 1) (RNase A).  
 GN RNASE1 OR RNS1.  
 OS Ovis aries (Sheep), and  
 OS Capra hircus (Goat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Caprinae; Ovis.  
 OX NCBI\_TaxID=9940, 9925;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Sheep, and C.hircus;  
 RX MEDLINE=96139017; PubMed=8587129;  
 RA Confalone E., Beintema J.J., Sasso M.P., Carsana A., Palmieri M.,  
 RA Vento M.T., Furia A.;  
 RT "Molecular evolution of genes encoding ribonucleases in ruminant  
 species.";  
 RL J. Mol. Evol. 41:850-858(1995).  
 RN [2]  
 RP SEQUENCE.  
 RC SPECIES=Sheep, and C.hircus; TISSUE=Pancreas;  
 RX MEDLINE=74309062; PubMed=4855010;  
 RA Wellling G.W., Scheffer A.J., Beintema J.J.;  
 RT "The primary structure of goat and sheep pancreatic ribonucleases.";  
 RL FEBS Lett. 41:58-61(1974).  
 RN [3]  
 RP SEQUENCE.  
 RC SPECIES=Sheep;  
 RX MEDLINE=74025520; PubMed=4356260;  
 RA Kobayashi R., Hirs C.H.W.;  
 RT "The amino acid sequence of ovine pancreatic ribonuclease A.";  
 RL J. Biol. Chem. 248:7833-7837(1973).  
 CC -|- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-  
 phosphates and 3'-phosphooligonucleotides ending in C-P or U-P  
 with 2',3'-cyclic phosphate intermediates.

```

CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: PANCREAS.
CC -1- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; S81741; AAB36135.1; -
CC EMBL; S81742; AAB36136.1; -
CC PIR; A00809; NRSH.
CC PIR; A00810; NRGT.
CC HSP; P00656; LSRN.
CC InterPro: IPR001427; RNaseA.
CC Pfam: PF00074; RNaseA; 1.
CC PRINTS; PR00794; RIBONUCLEASE.
CC ProDom; PD000535; RNaseA; 1.
CC SMART; SM00092; RNase_Pc; 1.
CC PROSITE; PS00127; RNASE_PANCREATIC; 1.
CC Hydrolase; Nuclease; Endonuclease; Glycoprotein.
CC FT DISULFID 26
CC FT DISULFID 40
CC FT DISULFID 58
CC FT DISULFID 58
CC FT ACT_SITE 12
CC FT ACT_SITE 12
CC FT ACT_SITE 41
CC FT ACT_SITE 119
CC FT CARBOHYD 34
CC FT CONFLICT 49
CC FT CONFLICT 103
CC FT CONFLICT 103
CC SEQUENCE 124 AA; 13707 MW; 3570EC1452BEDDFA CRC64;
CC -----
Query Match 2.88; Score 12; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.00012;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC -----
OY 22 AA AKAFERQHMDS 33
DB 4 AA AKAFERQHMDS 15
DB -----
RESULT 16
RNS_BOVIN STANDARD; PRT; 150 AA.
AC P00669;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ribonuclease, seminal precursor (EC 3.1.27.5) (Seminal RNase)
DE (S-RNase) (Ribonuclease BS-1).
DE SRN.
GN Bos taurus (Bovine).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE-90192098; PubMed-2315023;
RT Preuss K.D., Wagner S., Freudenstein J., Scheit K.H.;
RT "Cloning of cDNA encoding the complete precursor for bovine seminal
RT ribonuclease."
RL Nucleic Acids Res. 18:1057-1057(1990).
RN [2]
RN SEQUENCE FROM N.A.
RA Sasso M.P., Lombardi M., Confalone E., Carsana A., Palmieri M.,
RA Furia A.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [3]

```

```

RP SEQUENCE OF 27-150 FROM N.A.
RX MEDLINE-96139017; PubMed-8587129;
RA Confalone E., Beintema J.J., Sasso M.P., Carsana A., Palmieri M.,
RA Vento M.T., Furia A.;
RT "Molecular evolution of genes encoding ribonucleases in ruminant
RT species."
RL J. Mol. Evol. 41:850-858(1995).
RN [4]
RP SEQUENCE OF 27-150.
RA Suzuki H., Greco L., Parente A., Farina B., la Montagna R., Leone E.;
RT "Primary structure of seminal ribonuclease (RNase BS-1)."
RL Acta Vitaminol. Enzymol. 26:213-214(1972).
RN [5]
RP SEQUENCE OF 73-150 FROM N.A.
RX MEDLINE-86030265; PubMed-3840434;
RA Palmieri M., Carsana A., Furia A., Libonati M.;
RT "Sequence analysis of a cloned cDNA coding for bovine seminal
RT ribonuclease."
RL Eur. J. Biochem. 152:275-277(1985).
RN [6]
RP REVISION TO 43.
RX MEDLINE-83202042; PubMed-6846794;
RA Krietsch W.K.G., Simm F.C., Hertenberger B., Kuntz G.W.K., Wachter E.;
RT "Isolation of bovine seminal ribonuclease by affinity
RT chromatography."
RL Anal. Biochem. 128:213-216(1983).
RN [7]
RP INTERCHAIN DISULFIDE BONDS.
RX MEDLINE-74049879; PubMed-4761089;
RA di Donato A., D'Alessio G.;
RT "Interchain disulfide bridges in ribonuclease BS-1."
RL Biochem. Biophys. Res. Commun. 55:919-928(1973).
RN [8]
RP INTRACHAIN DISULFIDE BONDS.
RX MEDLINE-80130655; PubMed-534646;
RA di Donato A., D'Alessio G.;
RT "Intrachain disulfide bridges of bovine seminal ribonuclease."
RL Biochim. Biophys. Acta 579:303-313(1979).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE-84179398; PubMed-6673761;
RA Capasso S., Giordano F., Mattia C.A., Mazzarella L., Zagari A.;
RT "Refinement of the structure of bovine seminal ribonuclease."
RL Biopolymers 22:327-332(1983).
RN [10]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RA Mazzarella L., Capasso S., Demasi D., di Lorenzo G., Mattia C.A.,
RA Zagari A.;
RT "Bovine seminal ribonuclease: structure at 1.9-A resolution."
RL Acta Crystallogr. D 49:389-402(1993).
RN [11]
RP X-RAY CRYSTALLOGRAPHY (2.06 ANGSTROMS).
RX MEDLINE-99180403; PubMed-10082366;
RA Vitaliano L., Adinolfi S., Riccio A., Sica F., Zagari A.,
RA Mazzarella L.;
RT "Binding of a substrate analog to a domain swapping protein: X-ray
RT structure of the complex of bovine seminal ribonuclease with
RT uridylyl(2',5')adenosine."
RL Protein Sci. 7:1691-1699(1998).
RN [12]
RP MULTIPLE FORMS.
RX MEDLINE-82091855; PubMed-7317378;
RA di Donato A., D'Alessio G.;
RT "Heterogeneity of bovine seminal ribonuclease."
RL Biochemistry 20:7232-7237(1981).
RN [13]
RP ENZYMATIC ACTIVITY.
RX MEDLINE-73213326; PubMed-4664228;
RA D'Alessio G., Parente A., Guida C., Leone E.;
RT "Dimeric structure of seminal ribonuclease."
RL FEBS Lett. 27:285-288(1972).
RN [14]
RP REVIEW.

```

RX MEDLINE-91280661; PubMed-2057997;  
RA D'Alessio G., di Donato A., Parente A., Piccoli R.;  
RT "Seminal RNase: a unique member of the ribonuclease superfamily.";  
RL Trends Biochem. Sci. 16:104-106(1991).  
CC -1- FUNCTION: THIS ENZYME HYDROLYZES BOTH SINGLE- AND DOUBLE-STRANDED  
RNA.  
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-  
CC phosphates and 3'-phosphooligonucleotides ending in C-P or U-P  
CC with 2',3'-cyclic phosphate intermediates.  
CC -1- ENZYME REGULATION: ALLOSTERIC REGULATION BY BOTH SUBSTRATE AND  
CC REACTION PRODUCTS.  
CC -1- SUBUNIT: HOMODIMER; DISULFIDE-LINKED.  
CC -1- TISSUE SPECIFICITY: SEMINAL PLASMA. CAN REACH 3% OF THE PROTEIN  
CC CONTENT OF THIS FLUID.  
CC -1- MISCELLANEOUS: PROGRESSIVE DEAMINATION OF ASN-93 TRANSFORMS THE  
CC HOMODIMER (BETA- 2) INTO AND HETERODIMER (ALPHA-BETA) AND FINALLY  
CC A DOUBLY DEAMINATED DIMER (ALPHA-2).  
CC -1- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: X51337; CAA35716.1; -;  
CC EMBL: AJ000518; CAA04155.1; -;  
CC EMBL: S81747; AAB36140.1; -;  
CC EMBL: X03029; CAA26832.1; -;  
CC PIR: S08392; NRBO.  
CC PDB: 1BSR; 31-OCT-93.  
CC PDB: 11BA; 26-MAR-99.  
CC PDB: 11BG; 05-NOV-99.  
CC PDB: 1CQJ; 07-MAY-99.  
CC PDB: 1CGR; 07-MAY-99.  
CC InterPro: IPR001427; RNaseA.  
CC Pfam: PF00074; rnasea; 1.  
CC PRINTS: PR00794; RIBONUCLEASE.  
CC ProDom: PD000535; RNaseA; 1.  
CC SMART: SM00092; RNase\_Pc; 1.  
CC PROSITE: PS00127; RNASE\_PANCREATIC; 1.  
KW Hydrolase; Nuclease; Endonuclease; Signal; Allosteric enzyme;  
KW 3D-structure.  
FT SIGNAL 1 26  
FT CHAIN 27 150 RIBONUCLEASE, SEMINAL.  
FT ACT\_SITE 38 38 BY SIMILARITY.  
FT ACT\_SITE 67 67 BY SIMILARITY.  
FT ACT\_SITE 145 145 BY SIMILARITY.  
FT DISULFID 52 110  
FT DISULFID 66 121  
FT DISULFID 84 136  
FT DISULFID 91 98  
FT DISULFID 57 57 INTERCHAIN:  
FT DISULFID 58 58 INTERCHAIN.  
FT MOD\_RES 93 93 DEAMINATION (PROGRESSIVE).  
FT HELIX 30 38  
FT STRAND 39 39  
FT TURN 45 46  
FT TURN 48 58  
FT TURN 59 60  
FT STRAND 69 73  
FT HELIX 77 85  
FT STRAND 87 89  
FT TURN 92 93  
FT STRAND 98 100  
FT STRAND 105 112  
FT TURN 114 115  
FT STRAND 117 117  
FT TURN 118 119  
FT STRAND 120 120  
FT STRAND 123 130

FT STRAND 132 137  
FT TURN 138 141  
FT STRAND 142 149  
SQ SEQUENCE 150 AA; 16377 MW; F7A05C930FB83A83 CRC64;  
Query Match 2.8%; Score 12; DB 1; Length 150;  
Best Local Similarity 100.0%; Pred. No. 0.00015;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 22 AAKFERQHMD 33  
DB 30 AAKFERQHMD 41  
|||||||  
|||||||  
RESULT 17  
RNP\_ALCAA STANDARD; PRT; 124 AA.  
AC P00667;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE: Ribonuclease pancreatic (EC 3.1.27.5) (RNase A).  
GN RNASE1 OR RNS1.  
OS Alces alces alces (European moose) (Elk).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Cervioidea;  
OC Cervidae; Odocoileinae; Alces.  
OX NCBI\_TaxID=9853;  
RN [1]  
RP PARTIAL SEQUENCE.  
RC TISSUE=Pancreas;  
RX MEDLINE=76003215; PubMed=1157925;  
RA Leijenaar-Van den Berg G., Beintema J.J.;  
RT "The amino acid sequences of reindeer, moose and fallow deer  
RT pancreatic ribonucleases";  
RL FEBS Lett. 56:101-107(1975).  
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-  
CC phosphates and 3'-phosphooligonucleotides ending in C-P or U-P.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- TISSUE SPECIFICITY: PANCREAS.  
CC -1- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.  
DR PIR; C91418; NREKN.  
DR HSP; P00636; LSRN.  
DR InterPro: IPR001427; RNaseA.  
DR Pfam: PF00074; rnasea; 1.  
DR PRINTS: PR00794; RIBONUCLEASE.  
DR ProDom: PD000535; RNaseA; 1.  
DR SMART: SM00092; RNase\_Pc; 1.  
DR PROSITE: PS00127; RNASE\_PANCREATIC; 1.  
KW Hydrolase; Nuclease; Endonuclease; Glycoprotein.  
FT DISULFID 26 84 BY SIMILARITY.  
FT DISULFID 40 95 BY SIMILARITY.  
FT DISULFID 58 110 BY SIMILARITY.  
FT DISULFID 65 72 BY SIMILARITY.  
FT ACT\_SITE 12 12 BY SIMILARITY.  
FT ACT\_SITE 41 41 BY SIMILARITY.  
FT ACT\_SITE 119 119 BY SIMILARITY.  
FT CARBOHYD 34 34 N-LINKED (GLCNAC...).  
SQ SEQUENCE 124 AA; 13774 MW; 49F9C27F97DE03A7 CRC64;  
Query Match 2.6%; Score 11; DB 1; Length 124;  
Best Local Similarity 100.0%; Pred. No. 0.0013;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 22 AAKFERQHMD 32  
DB 4 AAKFERQHMD 14  
|||||||  
|||||||  
RESULT 18  
RNP\_AXIPR STANDARD; PRT; 124 AA.  
ID RNP\_AXIPR

AC P87351;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Ribonuclease pancreatic (EC 3.1.27.5) (RNase 1) (RNase A).  
GN RNASE1 OR RNS1.  
OS Axis porcinus (Hog deer).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Cervoidae;  
OC Cervidae; Cervinae; Axis.  
OX NCBI\_TaxID=57737;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98278842; PubMed=9611269;  
RA Breukelman H.J., van der Munnik N., Kleineidam R.G., Furia A.,  
RA Beintema J.J.;  
RT "Secretory ribonuclease genes and pseudogenes in true ruminants.";  
RL Gene 212:259-268(1998).  
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-  
CC phosphates and 3'-phosphooligonucleotides ending in C-P or U-P  
CC with 2',3'-cyclic phosphate intermediates.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- TISSUE SPECIFICITY: PANCREAS.  
CC -1- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; Y11669; CAA72367.1; -.  
DR HSP: P00656; 2RNS.  
DR InterPro: IPR001427; RNaseA.  
DR Pfam; PF00074; RNaseA; 1.  
DR PRINTS; PR00794; RIBONUCLEASE.  
DR PRODOM; PD000535; RNaseA; 1.  
DR SMART; SM00092; RNase\_Pc; 1.  
DR PROSITE; PS00127; RNASE\_PANCREATIC; 1.  
KW Hydrolase; Nuclease; Endonuclease.  
FT DISULFID 26 84 BY SIMILARITY.  
FT DISULFID 40 95 BY SIMILARITY.  
FT DISULFID 58 110 BY SIMILARITY.  
FT DISULFID 65 72 BY SIMILARITY.  
FT ACT\_SITE 12 12 BY SIMILARITY.  
FT ACT\_SITE 41 41 BY SIMILARITY.  
FT ACT\_SITE 119 119 BY SIMILARITY.  
FT ACT\_SITE 124 124 BY SIMILARITY.  
SQ SEQUENCE 124 AA; 13750 MW; B182A487553D3915 CRC64;  
  
Query Match 2.6%; Score 11; DB 1; Length 124;  
Best Local Similarity 100.0%; Pred. No. 0.0013;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 22 AA AKAFERQHMD 32  
IIIIIIIIII  
DB 4 AA AKAFERQHMD 14  
  
RESULT 19  
RNP\_CAPCA STANDARD; PRT; 124 AA.  
AC P00664;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Ribonuclease pancreatic (EC 3.1.27.5) (RNase 1) (RNase A).  
GN RNASE1 OR RNS1.  
OS Capreolus capreolus (Roe deer).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Cervoidae;  
OC Cervidae; Odocoileinae; Capreolus.

OX NCBI\_TaxID=9858;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Pancreas;  
RX MEDLINE=73253236; PubMed=4738402;  
RA Zwiers H., Scheffer A.J., Beintema J.J.;  
RT "Amino-acid sequences of red-deer and roe-deer pancreatic  
RT ribonucleases.";  
RL Eur. J. Biochem. 36:569-574(1973).  
RN [2]  
RP REVISIONS.  
RX MEDLINE=77112628; PubMed=836889;  
RA Oosterhuis S., Welling G.W., Gaastra W., Beintema J.J.;  
RT "Reinvestigation of the primary structures of red deer and roe deer  
RT pancreatic ribonuclease and proline sites in mammalian  
RT ribonucleases.";  
RL Biochim. Biophys. Acta 490:523-529(1977).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98278842; PubMed=9611269;  
RA Breukelman H.J., van der Munnik N., Kleineidam R.G., Furia A.,  
RA Beintema J.J.;  
RT "Secretory ribonuclease genes and pseudogenes in true ruminants.";  
RL Gene 212:259-268(1998).  
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-  
CC phosphates and 3'-phosphooligonucleotides ending in C-P or U-P  
CC with 2',3'-cyclic phosphate intermediates.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- TISSUE SPECIFICITY: PANCREAS.  
CC -1- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; Y11672; CAA72370.1; -.  
DR PIR; A90613; NRDEO.  
DR HSP: P00656; 1SRN.  
DR InterPro: IPR001427; RNaseA.  
DR Pfam; PF00074; RNaseA; 1.  
DR PRINTS; PR00794; RIBONUCLEASE.  
DR PRODOM; PD000535; RNaseA; 1.  
DR SMART; SM00092; RNase\_Pc; 1.  
DR PROSITE; PS00127; RNASE\_PANCREATIC; 1.  
KW Hydrolase; Nuclease; Endonuclease; Glycoprotein.  
FT DISULFID 26 84 BY SIMILARITY.  
FT DISULFID 40 95 BY SIMILARITY.  
FT DISULFID 58 110 BY SIMILARITY.  
FT DISULFID 65 72 BY SIMILARITY.  
FT ACT\_SITE 12 12 BY SIMILARITY.  
FT ACT\_SITE 41 41 BY SIMILARITY.  
FT ACT\_SITE 119 119 BY SIMILARITY.  
FT ACT\_SITE 124 124 BY SIMILARITY.  
FT CARBOHYD 34 34 N-LINKED (GLCNAC... ) (PARTIAL).  
SQ SEQUENCE 124 AA; 3836B2017625C613 CRC64;  
  
Query Match 2.6%; Score 11; DB 1; Length 124;  
Best Local Similarity 100.0%; Pred. No. 0.0013;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 22 AA AKAFERQHMD 32  
IIIIIIIIII  
DB 4 AA AKAFERQHMD 14  
  
RESULT 20  
RNP\_CEREL STANDARD; PRT; 124 AA.  
ID RNP\_CEREL  
AC P00663;  
DT 21-JUL-1986 (Rel. 01, Created)



DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DE Ribonuclease pancreatic (EC 3.1.27.5) (RNase 1) (RNase A).  
 GN RNASE1 OR RNS1.  
 OS Cervus elaphus (Red deer).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Cervoidae;  
 OC Cervidae; Cervinae; Cervus.  
 OX NCBI\_TaxID=9860;  
 RN [1]  
 RP PARTIAL SEQUENCE.  
 RC MEDLINE=76003215; PubMed=4738402;  
 RX MEDLINE=7323336; PubMed=4738402;  
 RA Zwiers H., Scheffer A.J., Beintema J.J.;  
 RT "Amino-acid sequences of red-deer and roe-deer pancreatic  
 ribonucleases.";  
 RL Eur. J. Biochem. 36:569-574(1973).  
 RN [2]  
 RP REVISIONS.  
 RX MEDLINE=77112528; PubMed=836889;  
 RA Oosterhuis S., Wellings G.W., Gastra W., Beintema J.J.;  
 RT "Reinvestigation of the primary structures of red deer and roe deer  
 pancreatic ribonuclease and proline sites in mammalian  
 ribonucleases.";  
 RL Biochim. Biophys. Acta 490:523-529(1977).  
 CC -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-  
 phosphates and 3'-phosphooligonucleotides ending in C-P or U-P  
 with 2',3'-cyclic phosphate intermediates.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: PANCREAS.  
 CC -!- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.  
 DR PIR; B90613; NRDER.  
 DR HSSP; P00656; ISRN.  
 DR InterPro; IPR001427; RNaseA.  
 DR Pfam; PF00074; rnaaseA; 1.  
 DR PRINTS; PR00794; RIBONUCLEASE.  
 DR ProDom; PD000535; RNaseA; 1.  
 DR SMART; SM00092; RNASE\_PC; 1.  
 DR PROSITE; PS00127; RNASE\_PANCREATIC; 1.  
 KW Hydrolase; Nuclease; Endonuclease.  
 FT DISULFID 26 84 BY SIMILARITY.  
 FT DISULFID 40 95 BY SIMILARITY.  
 FT DISULFID 58 110 BY SIMILARITY.  
 FT DISULFID 65 72 BY SIMILARITY.  
 FT ACT\_SITE 12 12 BY SIMILARITY.  
 FT ACT\_SITE 41 41 BY SIMILARITY.  
 FT ACT\_SITE 119 119 BY SIMILARITY.  
 SQ SEQUENCE 124 AA; 13777 MW; 330F86D52B3DD71B CRC64;  
  
 Query Match 2.6%; Score 11; DB 1; Length 124;  
 Best Local Similarity 100.0%; Pred. NO. 0.0013;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 22 AAKFERQHMD 32  
 Db 4 AAKFERQHMD 14  
  
 RESULT 21  
 RNP\_DAMDA  
 ID RNP\_DAMDA STANDARD; PRT; 124 AA.  
 AC P00665;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Ribonuclease pancreatic (EC 3.1.27.5) (RNase 1) (RNase A).  
 GN RNASE1 OR RNS1.  
 OS Dama dama (fallow deer) (Cervus dama).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Cervoidae;  
 OC Cervidae; Cervinae; Cervus.  
 OX NCBI\_TaxID=30532;  
 RN [1]

RP PARTIAL SEQUENCE.  
 RX MEDLINE=76003215; PubMed=1157925;  
 RA Leijenaar-Van den Berg G., Beintema J.J.;  
 RT "The amino acid sequences of reindeer, moose and fallow deer  
 pancreatic ribonucleases.";  
 RL FEBS Lett. 56:101-107(1975).  
 CC -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-  
 phosphates and 3'-phosphooligonucleotides ending in C-P or U-P  
 with 2',3'-cyclic phosphate intermediates.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: PANCREAS.  
 CC -!- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.  
 DR PIR; A91418; NRDEF.  
 DR HSSP; P00656; ISRN.  
 DR InterPro; IPR001427; RNaseA.  
 DR Pfam; PF00074; rnaaseA; 1.  
 DR PRINTS; PR00794; RIBONUCLEASE.  
 DR ProDom; PD000535; RNaseA; 1.  
 DR SMART; SM00092; RNASE\_PC; 1.  
 DR PROSITE; PS00127; RNASE\_PANCREATIC; 1.  
 KW Hydrolase; Nuclease; Endonuclease.  
 FT DISULFID 26 84 BY SIMILARITY.  
 FT DISULFID 40 95 BY SIMILARITY.  
 FT DISULFID 58 110 BY SIMILARITY.  
 FT DISULFID 65 72 BY SIMILARITY.  
 FT ACT\_SITE 12 12 BY SIMILARITY.  
 FT ACT\_SITE 41 41 BY SIMILARITY.  
 FT ACT\_SITE 119 119 BY SIMILARITY.  
 SQ SEQUENCE 124 AA; 13807 MW; A871BE3BFACE1827 CRC64;  
  
 Query Match 2.6%; Score 11; DB 1; Length 124;  
 Best Local Similarity 100.0%; Pred. NO. 0.0013;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 22 AAKFERQHMD 32  
 Db 4 AAKFERQHMD 14  
  
 RESULT 22  
 RNP\_RANTA  
 ID RNP\_RANTA STANDARD; PRT; 124 AA.  
 AC P00666;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Ribonuclease pancreatic (EC 3.1.27.5) (RNase 1) (RNase A).  
 GN RNASE1 OR RNS1.  
 OS Rangifer tarandus (Reindeer) (Caribou).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Cervoidae;  
 OC Cervidae; Odocoileinae; Rangifer.  
 OX NCBI\_TaxID=9870;  
 RN [1]  
 RP PARTIAL SEQUENCE.  
 RC TISSUE=Pancreas;  
 RX MEDLINE=76003215; PubMed=1157925;  
 RA Leijenaar-Van den Berg G., Beintema J.J.;  
 RT "The amino acid sequences of reindeer, moose and fallow deer  
 pancreatic ribonucleases.";  
 RL FEBS Lett. 56:101-107(1975).  
 CC -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-  
 phosphates and 3'-phosphooligonucleotides ending in C-P or U-P  
 with 2',3'-cyclic phosphate intermediates.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: PANCREAS.  
 CC -!- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.  
 DR PIR; B91418; NRDEF.  
 DR HSSP; P00656; ISRN.  
 DR InterPro; IPR001427; RNaseA.  
 DR Pfam; PF00074; rnaaseA; 1.  
 DR PRINTS; PR00794; RIBONUCLEASE.  
 DR ProDom; PD000535; RNaseA; 1.

```

DR SMART; SM00092; RNase_Pc; 1.
DR PROSITE; PS00127; RNASE_PANCREATIC; 1.
KW Hydrolase; Nuclease; Endonuclease.
FT DISULFID 26 84 BY SIMILARITY.
FT DISULFID 40 95 BY SIMILARITY.
FT DISULFID 58 110 BY SIMILARITY.
FT DISULFID 65 72 BY SIMILARITY.
FT ACT_SITE 12 12 BY SIMILARITY.
FT ACT_SITE 41 41 BY SIMILARITY.
FT ACT_SITE 119 119 BY SIMILARITY.
FT VARIANT 103 103
SQ SEQUENCE 124 AA; 13759 MW; CE568B8B9A87CD54 CRC64;

Query Match 2.6%; Score 11; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 AA AKFEROHMD 32
DB 4 AA AKFEROHMD 14

RESULT 23
RNP_CAMDR
ID RNP_CAMDR STANDARD; PRT; 124 AA.
AC P00670;
DT 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ribonuclease pancreatic (EC 3.1.27.5) (RNase 1) (RNase A).
GN RNASE1 OR RNS1.
OS Camelus dromedarius (Dromedary) (Arabic camel), and
OS Camelus bactrianus (Bactrian camel).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Camelus.
OX NCBI_TaxID=9838, 9837;
RN [1]
RP SEQUENCE.
RC SPECIES=C.dromedarius;
RX MEDLINE=76039472; PubMed=1167157;
RA Welling G.W., Groen G., Beintema J.J.;
RT "The amino acid sequence of dromedary pancreatic ribonuclease.";
RL Biochem. J. 147:505-511(1975).
RN [2]
RP MAJOR VARIANT, PARTIAL SEQUENCE, AND REVISIONS.
RC SPECIES=C.dromedarius;
RX MEDLINE=85204378; PubMed=3922790;
RA Beintema J.J.;
RT "Mammalian ribonucleases. The absence of a glycosylated Asn-Pro-Thr
sequence in horse ribonuclease and the presence of tryptophan at
position 39 in horse and dromedary ribonuclease.";
RL FEBS Lett. 185:115-120(1985).
RN [3]
RP PARTIAL SEQUENCE, AND VARIANT ALLELIC.
RC SPECIES=C.bactrianus;
RX MEDLINE=76277807; PubMed=962846;
RA Welling G.W., Mulder H., Beintema J.J.;
RT "Allelic polymorphism in arabian camel ribonuclease and the amino
acid sequence of bactrian camel ribonuclease.";
RL Biochem. Genet. 14:309-317(1976).
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-
phosphates and 3'-phosphooligonucleotides ending in C-P or U-P
with 2',3'-cyclic phosphate intermediates.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: PANCREAS.
CC -1- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.
DR PIR; A00815; NRCM.
DR PIR; A90229; NRCMM.
DR PIR; B90229; NRCMB.
DR HSP; P00656; ISRN.
DR InterPro; IPR001427; RNaseA.
DR Pfam; PF000074; rnasea; 1.
DR PRINTS; PR00794; RIBONUCLEASE.
DR ProDom; PD000535; RNaseA; 1.

```

---

```

DR SMART; SM00092; RNase_Pc; 1.
DR PROSITE; PS00127; RNASE_PANCREATIC; 1.
KW Hydrolase; Nuclease; Endonuclease.
FT DISULFID 26 84 BY SIMILARITY.
FT DISULFID 40 95 BY SIMILARITY.
FT DISULFID 58 110 BY SIMILARITY.
FT DISULFID 65 72 BY SIMILARITY.
FT ACT_SITE 12 12 BY SIMILARITY.
FT ACT_SITE 41 41 BY SIMILARITY.
FT ACT_SITE 119 119 BY SIMILARITY.
FT VARIANT 103 103
SQ SEQUENCE 124 AA; 13938 MW; 503ABA0BAAEFE92F CRC64;

Query Match 2.1%; Score 9; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KFERQHMD 33
DB 7 KFERQHMD 15

RESULT 24
RNP_GIRCA
ID RNP_GIRCA STANDARD; PRT; 124 AA.
AC P00662; Q29541; Q29534;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ribonuclease pancreatic (EC 3.1.27.5) (RNase 1) (RNase A).
GN RNASE1 OR RNS1.
OS Giraffa camelopardalis (Giraffe).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Giraffoidea;
OC Giraffidae; Giraffa.
OX NCBI_TaxID=9894;
RN [1]
RP SEQUENCE.
RA Gastra W.;
RL Thesis (1975), University of Groningen, Netherland.
RN [2]
RP SEQUENCE.
RX MEDLINE=74309061; PubMed=4855009;
RA Gastra W., Groen G., Welling G.W., Beintema J.J.;
RT "The primary structure of giraffe pancreatic ribonuclease.";
RL FEBS Lett. 41:227-232(1974).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=96139017; PubMed=8587129;
RA Confalone E., Beintema J.J., Sasso M.P., Carsana A., Palmieri M.,
RA Vento M.T., Furia A.;
RT "Molecular evolution of genes encoding ribonucleases in ruminant
species.";
RL J. Mol. Evol. 41:850-858(1995).
RN [4]
RP SEQUENCE OF 31-114 FROM N.A.
RX MEDLINE=93367815; PubMed=8360916;
RA Breukelman H.J., Beintema J.J., Confalone E., Costanzo C., Sasso M.P.,
RA Carsana A., Palmieri M., Furia A.;
RT "Sequences related to the ox pancreatic ribonuclease coding region in
the genomic DNA of mammalian species.";
RL J. Mol. Evol. 37:29-35(1993).
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-
phosphates and 3'-phosphooligonucleotides ending in C-P or U-P
with 2',3'-cyclic phosphate intermediates.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: PANCREAS.
CC -1- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; S81739; AAB36133.1; -;  
DR EMBL; S65127; AAB27932.1; -;  
DR PIR; A00811; NRGF.  
DR HSSP; P00656; 1SSB.  
DR InterPro: IPR001427; RNaseA.  
DR Pfam: PF00074; rNaseA; 1.  
DR PRINTS; PR00794; RIBONUCLEASE.  
DR ProDom; PD000535; RNaseA; 1.  
DR SMART; SM00092; RNase\_PC; 1.  
DR PROSITE; PS00127; RNASE\_PANCREATIC; 1.  
KW Hydrolase; Nuclease; Endonuclease; Glycoprotein.  
FT DISULFID 26 84 BY SIMILARITY.  
FT DISULFID 40 95 BY SIMILARITY.  
FT DISULFID 58 110 BY SIMILARITY.  
FT ACT\_SITE 12 12 BY SIMILARITY.  
FT ACT\_SITE 41 41 BY SIMILARITY.  
FT ACT\_SITE 119 119 BY SIMILARITY.  
FT CARBOHYD 34 34 N-LINKED (GLCNAC. . .).  
FT VARIANT 76 76 Y -> N.  
FT CONFLICT 28 28 E -> Q (IN REF. 1 AND 2).  
SQ SEQUENCE 124 AA; 13704 MW; CE97EBDC792612DA CRC64;

Query Match 2.1%; Score 9; DB 1; Length 124;  
Best Local Similarity 100.0%; Pred. No. 0.14;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 AAKFERQH 30  
DB 4 AAKFERQH 12

RESULT 25  
RNP\_MESAU  
ID RNP\_MESAU STANDARD; PRT; 124 AA.  
AC P00682;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Ribonuclease pancreatic (EC 3.1.27.5) (RNase 1) (RNase A).  
GN RNASE1 OR RNS1.  
OS Mesocricetus auratus (Golden hamster).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Cricetinae;  
OC Mesocricetus.  
OX NCBI\_TaxID=10036;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Pancreas;  
RA MEDLINE=80088445; PubMed=518928;  
RX Jekel P.A., Slips H.J., Lenstra J.A., Beintema J.J.;  
RT "The amino acid sequence of hamster pancreatic ribonuclease.";  
RL Biochimie 61:827-839(1979).  
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-phosphates and 3'-phosphooligonucleotides ending in C-P or U-P with 2',3'-cyclic phosphate intermediates.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- TISSUE SPECIFICITY: PANCREAS.  
CC -1- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.  
DR PIR; A00829; NRHY.  
DR HSSP; P00656; 1SRN.  
DR InterPro: IPR001427; RNaseA.  
DR Pfam; PF00074; rNaseA; 1.  
DR PRINTS; PR00794; RIBONUCLEASE.  
DR ProDom; PD000535; RNaseA; 1.  
DR SMART; SM00092; RNase\_PC; 1.  
DR PROSITE; PS00127; RNASE\_PANCREATIC; 1.  
KW Hydrolase; Nuclease; Endonuclease; Glycoprotein.

FT DISULFID 26 84 BY SIMILARITY.  
FT DISULFID 40 95 BY SIMILARITY.  
FT DISULFID 58 110 BY SIMILARITY.  
FT DISULFID 65 72 BY SIMILARITY.  
FT ACT\_SITE 12 12 BY SIMILARITY.  
FT ACT\_SITE 41 41 BY SIMILARITY.  
FT ACT\_SITE 119 119 BY SIMILARITY.  
FT CARBOHYD 34 34 N-LINKED (GLCNAC. . .).  
SQ SEQUENCE 124 AA; 13867 MW; 4E3C57F94E73C478 CRC64;

Query Match 2.1%; Score 9; DB 1; Length 124;  
Best Local Similarity 100.0%; Pred. No. 0.14;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KFERQHMS 33  
DB 7 KFERQHMS 15

RESULT 26  
RNP\_ONDZI  
ID RNP\_ONDZI STANDARD; PRT; 124 AA.  
AC P00681;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Ribonuclease pancreatic (EC 3.1.27.5) (RNase 1) (RNase A).  
GN RNASE1 OR RNS1.  
OS Ondatra zibethicus (Muskrat).  
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Arvicolinae;  
OC Ondatra.  
OX NCBI\_TaxID=10060;  
RN [1]  
RP SEQUENCE.

RX MEDLINE=76212474; PubMed=1279085;  
RA van Dijk H., Sloots B., van den Berg A., Gaastra W., Beintema J.J.;  
RT "The primary structure of muskrat pancreatic ribonuclease.";  
RL Int. J. Pept. Protein Res. 8:305-316(1976).  
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-phosphates and 3'-phosphooligonucleotides ending in C-P or U-P with 2',3'-cyclic phosphate intermediates.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- TISSUE SPECIFICITY: PANCREAS.  
CC -1- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.  
DR PIR; A00828; NROZ.  
DR HSSP; P00656; 1SRN.  
DR InterPro: IPR001427; RNaseA.  
DR Pfam; PF00074; rNaseA; 1.  
DR PRINTS; PR00794; RIBONUCLEASE.  
DR ProDom; PD000535; RNaseA; 1.  
DR SMART; SM00092; RNase\_PC; 1.  
DR PROSITE; PS00127; RNASE\_PANCREATIC; 1.  
KW Hydrolase; Nuclease; Endonuclease.  
FT DISULFID 26 84 BY SIMILARITY.  
FT DISULFID 40 95 BY SIMILARITY.  
FT DISULFID 58 110 BY SIMILARITY.  
FT DISULFID 65 72 BY SIMILARITY.  
FT ACT\_SITE 12 12 BY SIMILARITY.  
FT ACT\_SITE 41 41 BY SIMILARITY.  
FT ACT\_SITE 119 119 BY SIMILARITY.  
FT UNSURE 32 33  
FT UNSURE 75 76  
SQ SEQUENCE 124 AA; 13880 MW; A9E42DC90E9E7C04 CRC64;

Query Match 2.1%; Score 9; DB 1; Length 124;  
Best Local Similarity 100.0%; Pred. No. 0.14;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KFERQHMS 33  
DB 7 KFERQHMS 15

```

RESULT 27
RNP_HORSE
ID RNP_HORSE STANDARD; PRT; 128 AA.
AC P00674;
DT 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ribonuclease pancreatic (EC 3.1.27.5) (RNase 1) (RNase A).
GN RNASE1 OR RNS1.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN PRELIMINARY SEQUENCE.
RP MEDLINE=74302367; PubMed=4852291;
RX Scheffer A.J., Beintema J.J.;
RT "Horse pancreatic ribonuclease.";
RL Eur. J. Biochem. 46:221-233(1974).
RN [2]
RP REVISIONS.
RX MEDLINE=85204378; PubMed=3922790;
RA Beintema J.J.;
RT "Mammalian ribonucleases. The absence of a glycosylated Asn-Pro-Thr
sequence in horse ribonuclease and the presence of tryptophan at
position 39 in horse and dromedary ribonuclease.";
RL FEBS Lett. 185:115-120(1985).
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-
phosphates and 3'-phosphooligonucleotides ending in C-P or U-P
with 2',3'-cyclic phosphate intermediates.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: PANCREAS.
CC -1- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.
PIR: A00819; NRHO.
DR HSP; P00656; ISRN.
DR InterPro; IPR001427; RNaseA.
DR PRINTS; PR00794; RIBONUCLEASE.
DR PRODOM; PD000535; RNaseA; 1.
DR SMART; SM00092; RNase_Pc; 1.
DR PROSITE; PS00127; RNASE_PANCREATIC; 1.
KW Hydrolase; Nuclease; Endonuclease; Glycoprotein.
FT DISULFID 26 84 BY SIMILARITY.
FT DISULFID 40 95 BY SIMILARITY.
FT DISULFID 58 110 BY SIMILARITY.
FT DISULFID 65 72 BY SIMILARITY.
FT ACT_SITE 12 12 BY SIMILARITY.
FT ACT_SITE 41 41 BY SIMILARITY.
FT ACT_SITE 119 119 BY SIMILARITY.
FT CARBOHYD 34 34 N-LINKED (GLCNAC...).
FT VARIANT 98 98 G -> R.
FT SEQUENCE 128 AA; 14374 MW; A06727414097CIDD CRC64;
SQ
Query Match 2.1%; Score 9; DB 1; Length 128;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KFERQHMS 33
Db 7 KFERQHMS 15

RESULT 28
RNP_HYSCR
ID RNP_HYSCR STANDARD; PRT; 128 AA.
AC P04060;
DT 01-NOV-1986 (Rel. 03, Created)
DT 01-NOV-1986 (Rel. 03, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ribonuclease pancreatic (EC 3.1.27.5) (RNase 1) (RNase A).
GN RNASE1 OR RNS1.
OS Myocastor coypus (Coypu) (Nutria).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Myocastoridae;
OC Myocastor.
OX NCBI_TaxID=10157;
RN [1]
RP SEQUENCE.
RX MEDLINE=77065676; PubMed=999896;
RA van den Berg A., van den Hende-Timmer L., Beintema J.J.;
RT "Isolation, properties and primary structure of coypu and chinchilla
pancreatic ribonuclease.";
RL Biochim. Biophys. Acta 453:400-409(1976).
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-
phosphates and 3'-phosphooligonucleotides ending in C-P or U-P
with 2',3'-cyclic phosphate intermediates.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: PANCREAS.

```

CC -!- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.  
DR PIR: A00822; NRCU.  
DR HSSP: P00656; 1SRN.  
DR InterPro: IPR001427; RNaseA.  
DR Pfam: PF00074; rnaseA; 1.  
DR PRINTS: PR00794; RIBONUCLEASE.  
DR ProDom: PD000535; RNaseA; 1.  
DR SMART: SM00092; RNaseA; 1.  
DR PROSITE: PS00127; RNASE\_PANCREATIC; 1.  
KW Hydrolyase; Nuclease; Endonuclease; Glycoprotein.  
FT DISULFID 26 84 BY SIMILARITY.  
FT DISULFID 40 95 BY SIMILARITY.  
FT DISULFID 58 110 BY SIMILARITY.  
FT DISULFID 65 72 BY SIMILARITY.  
FT ACT\_SITE 12 12 BY SIMILARITY.  
FT ACT\_SITE 41 41 BY SIMILARITY.  
FT ACT\_SITE 119 119 BY SIMILARITY.  
FT CARBOHYD 34 34 N-LINKED (GLCNAC...).  
SQ SEQUENCE 128 AA; 14267 MW; 4EB924E52B445832 CRC64;

Query Match 2.1%; Score 9; DB 1; Length 128;  
Best Local Similarity 100.0%; Pred. No. 0.14;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KFERQHMS 33  
|||||  
DB 7 KFERQHMS 15

## RESULT 30

ID RNP\_CRILLO STANDARD; PRT; 130 AA.  
AC P24717;  
DT 01-MAR-1992 (Rel. 21, Created)  
DT 01-MAR-1992 (Rel. 21, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Ribonuclease pancreatic precursor (EC 3.1.27.5) (RNase 1) (RNase A) (Fragment).  
GN RNASE1 OR RNS1.  
OS Cricetus longicaudatus (Long-tailed hamster) (Chinese hamster).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.  
OX NCBI\_TaxID=10030;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-Ovary;  
RX MEDLINE=92158677; PubMed=1741299;  
RA Haug M., Schein C.H.;  
RT "The DNA sequences of the human and hamster secretory ribonucleases determined with the polymerase chain reaction (PCR).";  
RL Nucleic Acids Res. 20:612-612(1992).  
CC -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-phosphates and 3'-phosphooligonucleotides ending in C-P or U-P with 2',3'-cyclic phosphate intermediates.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- TISSUE SPECIFICITY: PANCREAS.  
CC -!- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X62945; CAA44717.1; -  
DR PIR: S22808; S22808.  
DR HSSP: P00656; 2RNS.  
DR InterPro: IPR001427; RNaseA.  
DR Pfam: PF00074; rnaseA; 1.  
DR ProDom: PD000535; RNaseA; 1.

DR SMART; SM00092; RNaseA; 1.  
DR PROSITE; PS00127; RNASE\_PANCREATIC; 1.  
KW Hydrolyase; Nuclease; Endonuclease; Glycoprotein; Signal.  
FT NON\_TER 1 1  
FT SIGNAL <1 6  
FT CHAIN 7 130 RIBONUCLEASE PANCREATIC.  
FT DISULFID 32 90 BY SIMILARITY.  
FT DISULFID 46 101 BY SIMILARITY.  
FT DISULFID 64 116 BY SIMILARITY.  
FT DISULFID 71 78 BY SIMILARITY.  
FT ACT\_SITE 18 18 BY SIMILARITY.  
FT ACT\_SITE 47 47 BY SIMILARITY.  
FT ACT\_SITE 125 125 BY SIMILARITY.  
FT CARBOHYD 40 40 N-LINKED (GLCNAC...). (PROBABLE).  
SQ SEQUENCE 130 AA; 14517 MW; 8D227FD6491B413D CRC64;

Query Match 2.1%; Score 9; DB 1; Length 130;  
Best Local Similarity 100.0%; Pred. No. 0.15;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KFERQHMS 33  
|||||  
DB 13 KFERQHMS 21

## RESULT 31

ID RNP\_CLEGL STANDARD; PRT; 148 AA.  
AC Q9WUR4;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Ribonuclease pancreatic precursor (EC 3.1.27.5) (RNase 1) (RNase A).  
GN RNASE1.  
OS Clethrionomys glareolus (Bank vole).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Arvicolinae; Clethrionomys.  
OX NCBI\_TaxID=51090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99439681; PubMed=10508551;  
RA Dubois J.-Y.F., Catzeffis F.M., Beintema J.J.;  
RT "The phylogenetic position of 'Acomyinae' (Rodentia, Mammalia) as sister group of a Murinae + Gerbillinae clade: evidence from the nuclear ribonuclease gene.";  
RL Mol. Phylogenet. Evol. 13:181-192(1999).  
CC -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-phosphates and 3'-phosphooligonucleotides ending in C-P or U-P with 2',3'-cyclic phosphate intermediates.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- TISSUE SPECIFICITY: PANCREAS.  
CC -!- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AJ005769; CAB41470.1; -  
DR HSSP: P00656; 1SRN.  
DR InterPro: IPR001427; RNaseA.  
DR Pfam: PF00074; rnaseA; 1.  
DR PRINTS: PR00794; RIBONUCLEASE.  
DR ProDom: PD000535; RNaseA; 1.  
DR SMART; SM00092; RNaseA; 1.  
DR PROSITE; PS00127; RNASE\_PANCREATIC; 1.  
KW Hydrolyase; Nuclease; Endonuclease; Glycoprotein; Signal.  
FT SIGNAL 1 25 POTENTIAL.  
FT CHAIN 26 148 RIBONUCLEASE PANCREATIC.

```
FT DISULFID 50 108 BY SIMILARITY..
FT DISULFID 62 119 BY SIMILARITY.
FT DISULFID 84 134 BY SIMILARITY.
FT DISULFID 89 96 BY SIMILARITY.
FT ACT_SITE 37 37 BY SIMILARITY.
FT ACT_SITE 65 65 BY SIMILARITY.
FT ACT_SITE 143 143 BY SIMILARITY.
FT CARBOHYD 58 58 N-LINKED (GLCNAC.. ) (POTENTIAL).
FT CARBOHYD 86 86 N-LINKED (GLCNAC.. ) (POTENTIAL).
SQ SEQUENCE 148 AA; 16471 MW; 028BD249C2DDB563 CRC64;

Query Match 2.1%; Score 9; DB 1; Length 148;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KFERQHMS 33
DB 32 KFERQHMS 40
|||||

RESULT 32
RNP_GERNI STANDARD; PRT; 148 AA.
AC Q9WU52:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ribonuclease pancreatic precursor (EC 3.1.27.5) (RNase A).
GN RNASE1.
OS Gerbillus nigeriae (Nigerian gerbil).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Gerbillinae;
OC Gerbillus.
OC NCBI_TaxID=39472;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99439681; PubMed=10508551;
RA Dubois J.-Y.F., Catzeffis F.M., Beintema J.J.:
RT "The phylogenetic position of 'Acomyinae' (Rodentia, Mammalia) as
RT sister group of a Murinae + Gerbillinae clade: evidence from the
RT nuclear ribonuclease gene."
RL Mol. Phylogenet. Evol. 13:181-192(1999).
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-
CC phosphates and 3'-phosphooligonucleotides ending in C-P or U-P
CC with 2',3'-cyclic phosphate intermediates.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: PANCREAS.
CC -1- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ005774; CAB41479.1;
CC HSSP; P00656; ISRN.
CC InterPro: IPR001427; RNaseA.
CC Pfam: PF00074; rnasea; 1.
CC PRINTS; PR00794; RIBONUCLEASE.
CC ProDom; PD000535; RNaseA; 1.
CC SMART; SM00092; RNase_Pc; 1.
CC PROSITE; PS00127; RNASE_PANCREATIC; 1.
CC KWL Hydrolase; Nuclease; Endonuclease; Signal.
CC SIGNAL 1 25 POTENTIAL.
CC CHAIN 26 148 RIBONUCLEASE PANCREATIC.
CC DISULFID 50 108 BY SIMILARITY.
CC DISULFID 64 119 BY SIMILARITY.
CC DISULFID 82 134 BY SIMILARITY.
CC DISULFID 89 96 BY SIMILARITY.
CC ACT_SITE 37 37 BY SIMILARITY.
CC ACT_SITE 65 65 BY SIMILARITY.
CC ACT_SITE 143 143 BY SIMILARITY.
SQ SEQUENCE 148 AA; 16331 MW; 264FE2DDEC351F9F CRC64;

Query Match 2.1%; Score 9; DB 1; Length 148;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
FT ACT_SITE 65 65 BY SIMILARITY.
FT ACT_SITE 143 143 BY SIMILARITY.
SQ SEQUENCE 148 AA; 16348 MW; 154FF5AE58C6B9A CRC64;

Query Match 2.1%; Score 9; DB 1; Length 148;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KFERQHMS 33
DB 32 KFERQHMS 40
|||||

RESULT 33
RNP_TATKG STANDARD; PRT; 148 AA.
AC Q9WU54:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ribonuclease pancreatic precursor (EC 3.1.27.5) (RNase A).
GN RNASE1.
OS Tatera kempii gambiana (Kemp's gerbil).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Gerbillinae;
OC Tatera.
OC NCBI_TaxID=41264;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99439681; PubMed=10508551;
RA Dubois J.-Y.F., Catzeffis F.M., Beintema J.J.:
RT "The phylogenetic position of 'Acomyinae' (Rodentia, Mammalia) as
RT sister group of a Murinae + Gerbillinae clade: evidence from the
RT nuclear ribonuclease gene."
RL Mol. Phylogenet. Evol. 13:181-192(1999).
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-
CC phosphates and 3'-phosphooligonucleotides ending in C-P or U-P
CC with 2',3'-cyclic phosphate intermediates.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: PANCREAS.
CC -1- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ005775; CAB41485.1;
CC HSSP; P00656; ISSA.
CC InterPro: IPR001427; RNaseA.
CC Pfam: PF00074; rnasea; 1.
CC PRINTS; PR00794; RIBONUCLEASE.
CC ProDom; PD000535; RNaseA; 1.
CC SMART; SM00092; RNase_Pc; 1.
CC PROSITE; PS00127; RNASE_PANCREATIC; 1.
CC KWL Hydrolase; Nuclease; Endonuclease; Signal.
CC SIGNAL 1 25 POTENTIAL.
CC CHAIN 26 148 RIBONUCLEASE PANCREATIC.
CC DISULFID 50 108 BY SIMILARITY.
CC DISULFID 64 119 BY SIMILARITY.
CC DISULFID 82 134 BY SIMILARITY.
CC DISULFID 89 96 BY SIMILARITY.
CC ACT_SITE 37 37 BY SIMILARITY.
CC ACT_SITE 65 65 BY SIMILARITY.
CC ACT_SITE 143 143 BY SIMILARITY.
SQ SEQUENCE 148 AA; 16331 MW; 264FE2DDEC351F9F CRC64;

Query Match 2.1%; Score 9; DB 1; Length 148;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 25 KFERQHMS 33
DB 32 KFERQHMS 40

RESULT 34
RNP_ACOCA
ID RNP_ACOCA STANDARD; PRT; 149 AA.
AC Q9WTT5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ribonuclease pancreatic precursor (EC 3.1.27.5) (RNase 1) (RNase A).
GN RNASE1.
OS Acomys cahirinus (Egyptian spiny mouse).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Acomys.
OX NCBI_TaxID=10068;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99439681; PubMed=10508551;
RA Dubois J.-Y.F., Catzeflis F.M., Beintema J.J.;
RT "The phylogenetic position of 'Acomyinae' (Rodentia, Mammalia) as
RT sister group of a Murinae + Gerbillinae clade: evidence from the
RT nuclear ribonuclease gene.";
RL Mol. Phylogenet. Evol. 13:181-192(1999).
CC -|- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-
CC phosphates and 3'-phosphooligonucleotides ending in C-P or U-P
CC with 2',3'-cyclic phosphate intermediates.
CC -|- SUBCELLULAR LOCATION: Secreted.
CC -|- TISSUE SPECIFICITY: PANCREAS.
CC -|- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AJ005772; CAB41464.1;
CC HSP: P00656; ISRN.
CC InterPro: IPR001427; RNaseA.
CC Pfam: PF00074; RNaseA.
CC PRINTS: PR00794; RIBONUCLEASE.
CC ProDom: PD000535; RNaseA.
CC SMART: SM00092; RNase_PC.
CC PROSITE: PS00127; RNASE_PANCREATIC; 1.
CC Hydrolase; Nuclease; Endonuclease; Glycoprotein; Signal.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 149 RIBONUCLEASE PANCREATIC.
FT DISULFID 51 109 BY SIMILARITY.
FT DISULFID 65 120 BY SIMILARITY.
FT DISULFID 83 135 BY SIMILARITY.
FT DISULFID 90 97 BY SIMILARITY.
FT ACT_SITE 37 37 BY SIMILARITY.
FT ACT_SITE 66 66 BY SIMILARITY.
FT ACT_SITE 144 144 BY SIMILARITY.
FT CARBOHYD 59 59 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 87 87 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 149 AA; 16613 MW; 9A25B5C11040456C CRC64;

Query Match 2.1%; Score 9; DB 1; Length 149;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KFERQHMS 33
DB 32 KFERQHMS 40

RESULT 35
RNP_MUSPA
ID RNP_MUSPA STANDARD; PRT; 149 AA.
AC Q90YX3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Ribonuclease pancreatic precursor (EC 3.1.27.5) (RNase 1) (RNase A).
GN RNASE1.
OS Akodon jelskii.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Sigmodontinae;
OX NCBI_TaxID=10079;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99439681; PubMed=10508551;
RA Dubois J.-Y.F., Catzeflis F.M., Beintema J.J.;
RT "The phylogenetic position of 'Acomyinae' (Rodentia, Mammalia) as
RT sister group of a Murinae + Gerbillinae clade: evidence from the
RT nuclear ribonuclease gene.";
RL Mol. Phylogenet. Evol. 13:181-192(1999).
CC -|- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-
CC phosphates and 3'-phosphooligonucleotides ending in C-P or U-P
CC with 2',3'-cyclic phosphate intermediates.
CC -|- SUBCELLULAR LOCATION: Secreted.
CC -|- TISSUE SPECIFICITY: PANCREAS.
CC -|- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AJ005771; CAB41465.1;
CC HSP: P00656; 2RNS.
CC InterPro: IPR001427; RNaseA.
CC Pfam: PF00074; RNaseA.
CC PRINTS: PR00794; RIBONUCLEASE.
CC ProDom: PD000535; RNaseA.
CC SMART: SM00092; RNase_PC.
CC PROSITE: PS00127; RNASE_PANCREATIC; 1.
CC Hydrolase; Nuclease; Endonuclease; Glycoprotein; Signal.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 149 RIBONUCLEASE PANCREATIC.
FT DISULFID 51 109 BY SIMILARITY.
FT DISULFID 65 120 BY SIMILARITY.
FT DISULFID 83 135 BY SIMILARITY.
FT DISULFID 90 97 BY SIMILARITY.
FT ACT_SITE 37 37 BY SIMILARITY.
FT ACT_SITE 66 66 BY SIMILARITY.
FT ACT_SITE 144 144 BY SIMILARITY.
FT CARBOHYD 59 59 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 87 87 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 149 AA; 16611 MW; 20F9CBE9089ABBD3 CRC64;

Query Match 2.1%; Score 9; DB 1; Length 149;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KFERQHMS 33
DB 32 KFERQHMS 40

RESULT 36
RNP_MUSPA
ID RNP_MUSPA STANDARD; PRT; 149 AA.
AC Q90YX3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)

```

15-JUN-2002 (Rel. 41, Last annotation update)  
Ribonuclease pancreatic precursor (EC 3.1.27.5) (RNase 1) (RNase A).  
RNASE1.  
Mus pahari (Shrew mouse).  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
NCBI\_TaxID=10093;  
[1]  
SEQUENCE FROM N.A.  
MEDLINE=99439681; PubMed=10508551;  
Dubois J.-Y.F., Catzeffis F.M., Beintema J.J.;  
"The phylogenetic position of 'Acomyinae' (Rodentia, Mammalia) as  
sister group of a Murinae + Gerbillinae clade: evidence from the  
nuclear ribonuclease gene.";  
Mol. Phylogenet. Evol. 13:181-192(1999).  
-1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-  
phosphates and 3'-phosphooligonucleotides ending in C-P or U-P  
with 2',3'-cyclic phosphate intermediates.  
-1- SUBCELLULAR LOCATION: Secreted.  
-1- TISSUE SPECIFICITY: PANCREAS.  
-1- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; AJ238699; CAB60000.1;  
HSP; P00656; 1SSA.  
InterPro: IPR001427; RNaseA.  
PRINTS; PR00794; RIBONUCLEASE.  
PRODOM; PD000535; RNaseA; 1.  
SMART; SM00092; RNase\_Pc; 1.  
PROSITE; PS00127; RNASE\_PANCREATIC; 1.  
Hydrolase; Nuclease; Endonuclease; Glycoprotein; Signal.  
FT CHAIN 1 25  
FT DISULFID 26 149 RIBONUCLEASE PANCREATIC.  
FT DISULFID 51 109 BY SIMILARITY.  
FT DISULFID 65 120 BY SIMILARITY.  
FT DISULFID 83 135 BY SIMILARITY.  
FT DISULFID 90 97 BY SIMILARITY.  
FT ACT\_SITE 37 37 BY SIMILARITY.  
FT ACT\_SITE 66 66 BY SIMILARITY.  
FT ACT\_SITE 144 144 BY SIMILARITY.  
FT CARBOHYD 87 87 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT SEQUENCE 149 AA; 16691 MW; 47905C8E0B3B54B4 CRC64;  
Query Match 2.18; Score 9; DB 1; Length 149;  
Best Local Similarity 100.0%; Pred. No. 0.17;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 25 KFERQHMD 33  
Db 32 KFERQHMD 40  
RESULT 37  
RNP\_URARU STANDARD; PRT; 149 AA.  
AC QWUX6;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Ribonuclease pancreatic precursor (EC 3.1.27.5) (RNase 1) (RNase A).  
GN RNASE1.  
OS Uranomys ruddi (White-bellied brush-furred rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
OC Uranomys.

NCBI\_TaxID=41272;  
[1]  
SEQUENCE FROM N.A.  
MEDLINE=99439681; PubMed=10508551;  
Dubois J.-Y.F., Catzeffis F.M., Beintema J.J.;  
"The phylogenetic position of 'Acomyinae' (Rodentia, Mammalia) as  
sister group of a Murinae + Gerbillinae clade: evidence from the  
nuclear ribonuclease gene.";  
Mol. Phylogenet. Evol. 13:181-192(1999).  
-1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-  
phosphates and 3'-phosphooligonucleotides ending in C-P or U-P  
with 2',3'-cyclic phosphate intermediates.  
-1- SUBCELLULAR LOCATION: Secreted.  
-1- TISSUE SPECIFICITY: PANCREAS.  
-1- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; AJ005773; CAB41486.1;  
HSP; P00656; 1SRN.  
InterPro: IPR001427; RNaseA.  
PRINTS; PR00794; RIBONUCLEASE.  
PRODOM; PD000535; RNaseA; 1.  
SMART; SM00092; RNase\_Pc; 1.  
PROSITE; PS00127; RNASE\_PANCREATIC; 1.  
Hydrolase; Nuclease; Endonuclease; Signal.  
FT CHAIN 1 25  
FT DISULFID 26 149 RIBONUCLEASE PANCREATIC.  
FT DISULFID 51 109 BY SIMILARITY.  
FT DISULFID 65 120 BY SIMILARITY.  
FT DISULFID 83 135 BY SIMILARITY.  
FT DISULFID 90 97 BY SIMILARITY.  
FT ACT\_SITE 37 37 BY SIMILARITY.  
FT ACT\_SITE 66 66 BY SIMILARITY.  
FT ACT\_SITE 144 144 BY SIMILARITY.  
FT SEQUENCE 149 AA; 16798 MW; 0531207ECBF5C1EE CRC64;  
Query Match 2.18; Score 9; DB 1; Length 149;  
Best Local Similarity 100.0%; Pred. No. 0.17;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 25 KFERQHMD 33  
Db 32 KFERQHMD 40  
RESULT 38  
HPN\_HELPY STANDARD; PRT; 59 AA.  
AC Q48251;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Histidine-rich, metal binding polypeptide.  
GN HPN OR HP1427 OR JHP1320.  
OS Helicobacter pylori (Campylobacter pylori), and  
OS Helicobacter pylori J99 (Campylobacter pylori J99).  
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
OC Helicobacter.  
OC NCBI\_TaxID=210, 85963;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-18 AND 47-59.  
RC STRAIN=LEU;  
RX MEDLINE=95310028; PubMed=7790085;  
Gilbert J.V., Ramakrishna J., Sunderman F.W. Jr., Wright A.,  
Plaut A.G.;





```

CC DIRECTLY PROMOTES THE ACCUMULATION OF EOSINOPHILS (A PROMINENT
CC FEATURE OF ALLERGIC INFLAMMATORY REACTIONS), BUT NOT
CC LYMPHOCYTES, MACROPHAGES OR NEUTROPHILS.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: EXPRESSED CONSTITUTIVELY IN THE THYMUS.
CC EXPRESSION INDUCIBLE IN THE LUNG (TYPE I-ALVEOLAR EPITHELIAL
CC CELLS), INTESTINE, HEART, SPLEEN, KIDNEY.
CC -!- INDUCTION: BY INTERFERON GAMMA AND LIPOPOLYSACCHARIDE (LPS).
CC -!- SIMILARITY: BELONGS TO THE INTERCRINE BETA FAMILY (SMALL CYTOKINE
CC C-C) (CHEMOKINE CC).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U26426; AAC52256.1; -;
CC EMBL; U40672; AAA99776.1; -;
CC HSPF; P51671; LEOT.
CC MGD; MGI:103576; Scyall.
CC InterPro; IPR000827; CC_chemkine_sml.
CC InterPro; IPR001811; Chemokine_IL8.
CC Pfam; PF00048; IL8; 1.
CC SMART; SM00199; SCY; 1.
CC PROSITE; PS00472; SMALL_CYTOKINES_CC; 1.
KW Eosinophil; Cytokine; Chemotaxis; Glycoprotein; Signal;
KW Inflammatory response.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 97 EOTAXIN.
FT DISULFID 32 57 BY SIMILARITY.
FT DISULFID 33 73 BY SIMILARITY.
FT CARBOHYD 94 94 O-LINKED (GALNAC...) (BY SIMILARITY).
SQ SEQUENCE 97 AA; 10893 MW; 36C9812107FC6CA7 CRC64;

Query Match 1.9%; Score 8; DB 1; Length 97;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 152 NTLKSYK 159
Db 43 NTLKSYK 50
|||||||

Search completed: May 12, 2003, 10:26:40
Job time : 42 secs

```

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 10:23:33 ; Search time 35 Seconds  
(without alignments)  
2537.323 Million cell updates/sec

Title: US-10-057-531A-2

Perfect score: 431

Sequence: 1 MHHHHHHHSLVPRSGMKE.....TCECTKPSYPLFDGIFCSS 431

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

SPTREMBL\_21.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_rvirus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	374	86.8	1720	5	Q25922 plasmodium
2	344	79.8	652	5	Q25923 plasmodium
3	293	68.0	373	5	Q25723 plasmodium
4	231	53.6	373	5	O43995 plasmodium
5	231	53.6	373	5	O43996 plasmodium
6	231	53.6	373	5	Q25724 plasmodium
7	192	44.5	373	5	Q25722 plasmodium
8	190	44.1	570	5	O9TYG2 plasmodium
9	190	44.1	570	5	Q25968 plasmodium
10	189	43.9	373	5	Q25721 plasmodium
11	165	38.3	373	5	Q25727 plasmodium
12	165	38.3	373	5	Q25728 plasmodium
13	145	33.6	372	5	O43997 plasmodium
14	145	33.6	372	5	Q25725 plasmodium
15	145	33.6	569	5	Q25978 plasmodium
16	145	33.6	569	5	Q25983 plasmodium

Q25924 plasmodium  
Q25970 plasmodium  
Q25980 plasmodium  
Q25982 plasmodium  
Q25717 plasmodium  
Q25719 plasmodium  
Q25720 plasmodium  
Q25726 plasmodium  
Q9TVG8 plasmodium  
Q25968 plasmodium  
Q25974 plasmodium  
Q25975 plasmodium  
Q25977 plasmodium  
Q25979 plasmodium  
Q9NHX1 plasmodium  
Q9TZT5 plasmodium  
Q9TZT4 plasmodium  
Q25718 plasmodium  
Q25967 plasmodium  
Q9TYE6 plasmodium  
Q964N3 plasmodium  
Q964N3 plasmodium  
Q9BJU5 plasmodium  
Q9BJU4 plasmodium  
Q9BJU3 plasmodium  
Q9BJU2 plasmodium  
Q9BJU1 plasmodium  
Q9UAI8 plasmodium  
Q9BJS7 plasmodium  
Q9BJS6 plasmodium  
Q964N2 plasmodium  
Q9TYE4 plasmodium  
Q25972 plasmodium  
Q25981 plasmodium  
Q25981 plasmodium  
Q9BJU6 plasmodium  
Q9BJS8 plasmodium  
Q9BJV6 plasmodium  
Q9BJV5 plasmodium  
Q9BJV4 plasmodium  
Q9BJV3 plasmodium  
Q9BJV2 plasmodium  
Q9BJV1 plasmodium  
Q9BJV0 plasmodium  
Q9BJU9 plasmodium  
Q9BJU8 plasmodium  
Q9BJU7 plasmodium  
Q9BJU0 plasmodium  
Q9BJT9 plasmodium  
Q9BJT8 plasmodium  
Q9BJT7 plasmodium  
Q9BJT6 plasmodium  
Q9BJT5 plasmodium  
Q9BJT4 plasmodium  
Q9BJT3 plasmodium  
Q9BJT2 plasmodium  
Q9BJT1 plasmodium  
Q9BJT0 plasmodium  
Q9BJS9 plasmodium  
Q9TYE5 plasmodium  
Q9TYE3 plasmodium  
Q9TYE7 plasmodium  
Q9UAI6 plasmodium  
Q9UAI7 plasmodium  
Q9UB87 plasmodium  
Q9BMG8 plasmodium  
Q25999 plasmodium  
Q25984 plasmodium  
Q25966 plasmodium  
Q9TYG1 plasmodium  
Q25976 plasmodium  
Q25971 plasmodium  
Q25973 plasmodium  
Q36012 escherichia  
O55244 escherichia

90	15	3.5	124	6	Q9T5F2	Q9tsf2 bos taurus	163	7	1.6	134	8	Q9GDG5	Q9gdg5 malus ioens
91	15	3.5	674	5	Q9GRE8	Q9gre8 toxoplasma	164	7	1.6	134	8	Q9GDG7	Q9gdg7 malus yunna
92	12	2.8	21	6	Q9TRQ8	Q9trq8 bos taurus	165	7	1.6	135	8	Q9GDG2	Q9gdg2 malus fusca
93	12	2.8	124	6	Q9M235	Q9m235 bubalus bub	166	7	1.6	136	4	Q9NXL9	Q9nxl9 homo sapien
94	12	2.8	144	6	Q9BH14	Q9bh14 antilocapra	167	7	1.6	136	4	Q9HCV5	Q9hcv5 homo sapien
95	9	2.1	116	6	Q9TVB7	Q9tvb7 camelus dro	168	7	1.6	136	8	Q9GIF1	Q9glf1 malus sieve
96	9	2.1	134	6	Q9BDB9	Q9bdb9 tragulus ja	169	7	1.6	136	8	Q9GIF0	Q9glf0 malus domes
97	9	2.1	168	4	Q16256	Q16256 homo sapien	170	7	1.6	136	8	Q9GDH0	Q9gdh0 malus flore
98	8	1.9	59	16	Q932C2	Q932c2 staphylococ	171	7	1.6	136	8	Q9GDG2	Q9gdg2 malus orien
99	8	1.9	116	6	Q9XSQ3	Q9xsq3 lama guanac	172	7	1.6	136	8	Q9GBG1	Q9gdb1 malus pratt
100	8	1.9	124	6	Q95NE6	Q95ne6 bubalus bub	173	7	1.6	136	8	Q9GDF1	Q9gdf1 malus coron
101	8	1.9	142	6	Q9BEC3	Q9bec3 tragulus ja	174	7	1.6	136	8	Q9GDE9	Q9gde9 pyrus salic
102	8	1.9	152	11	Q8VD89	Q8vdb9 rattus norv	175	7	1.6	140	17	Q9VAC0	Q9vac0 aeropyrum p
103	8	1.9	158	10	Q9SYL3	Q9sy13 arabidopsis	176	7	1.6	141	8	Q9GDF4	Q9gdf4 malus trilo
104	8	1.9	226	11	Q9QZD2	Q9qzd2 rattus norv	177	7	1.6	141	8	Q9GDF0	Q9gdf0 pyrus elaea
105	8	1.9	238	17	Q973R9	Q973r9 sulfobus	178	7	1.6	141	8	Q9GDE8	Q9gde8 docynia del
106	8	1.9	278	16	P74248	P74248 synechocyst	179	7	1.6	141	8	Q9GDE7	Q9gde7 cydonia obl
107	8	1.9	288	16	Q92LM2	Q92lm2 rhizobium m	180	7	1.6	151	8	Q9GBE6	Q9gbe6 malus hupeh
108	8	1.9	313	12	Q9YVU1	Q9yvul melanoplus	181	7	1.6	155	8	Q9GDH3	Q9gdh3 malus domes
109	8	1.9	362	10	Q9ATS8	Q9ats8 cymbopogon	182	7	1.6	155	8	Q9GDH1	Q9gdh1 malus bacca
110	8	1.9	365	16	Q9JQM5	Q9jqm5 neisseria m	183	7	1.6	155	8	Q9GDB8	Q9gdb8 malus fusca
111	8	1.9	366	10	Q9ATS4	Q9ats4 capillipedi	184	7	1.6	155	8	Q9GDF3	Q9gdf3 malus tscho
112	8	1.9	366	10	Q9ATS0	Q9ats0 bothriochlo	185	7	1.6	158	2	Q9E2I3	Q9ezi3 borrelia bu
113	8	1.9	379	16	Q9CFB0	Q9cfb0 lactococcus	186	7	1.6	158	5	Q94577	Q94577 helicoidari
114	8	1.9	416	5	Q96144	Q96144 plasmodium	187	7	1.6	162	17	Q95981	Q95981 pyrococcus
115	8	1.9	431	5	Q9VFP2	Q9vfp2 drosophila	188	7	1.6	168	17	Q9YBR2	Q9ybr2 aeropyrum p
116	8	1.9	503	8	Q8WJQ5	Q8wjq5 gillenia st	189	7	1.6	173	16	Q9PR48	Q9pr48 ureaplasma
117	8	1.9	516	13	Q42186	Q42186 brachydanio	190	7	1.6	176	2	Q9F933	Q9f933 borrelia bu
118	8	1.9	517	13	Q90YL6	Q90yl6 brachydanio	191	7	1.6	176	5	P90669	P90669 aplysia cal
119	8	1.9	560	10	Q94DW0	Q94dw0 oryza sativ	192	7	1.6	178	10	Q9LTP3	Q9ltp3 arabidopsis
120	8	1.9	632	9	Q9F2Y9	Q9fzy9 bacterioph	193	7	1.6	185	5	Q00837	Q00837 leishmania
121	8	1.9	636	9	Q8SDT4	Q8sdt4 bacterioph	194	7	1.6	190	17	Q9YAR2	Q9yar2 aeropyrum p
122	8	1.9	636	16	Q931W8	Q931w8 staphylococ	195	7	1.6	194	15	Q89968	Q89968 chimpanzee
123	8	1.9	678	5	Q94736	Q94736 stomoxys ca	196	7	1.6	194	15	Q9YVC7	Q9yvc7 chimpanzee
124	8	1.9	815	10	Q9FZ65	Q9fz65 arabidopsis	197	7	1.6	194	15	Q89973	Q89973 chimpanzee
125	8	1.9	923	13	Q90ZT7	Q90zt7 xenopus lae	198	7	1.6	194	15	Q9YVC6	Q9yvc6 chimpanzee
126	8	1.9	1281	13	Q918L4	Q918l4 brachydanio	199	7	1.6	194	15	Q9YVB8	Q9yvb8 chimpanzee
127	8	1.9	3309	5	Q9GUP2	Q9gup2 caenorhabdi	200	7	1.6	195	16	Q97KX0	Q97kx0 clostridium
128	7	1.6	44	12	Q65825	Q65825 unidentified	201	7	1.6	203	5	Q17346	Q17346 caenorhabdi
129	7	1.6	83	5	Q20689	Q20689 caenorhabdi	202	7	1.6	207	10	Q04307	Q04307 arabidopsis
130	7	1.6	88	16	P96637	P96637 bacillus su	203	7	1.6	209	16	Q45593	Q45593 bacillus su
131	7	1.6	90	17	Q9HJR9	Q9hjr9 thermoplas	204	7	1.6	210	16	Q07036	Q07036 mycobacteri
132	7	1.6	91	9	Q9AYU7	Q9ayu7 bacterioph	205	7	1.6	211	10	Q9M1F8	Q9m1f8 arabidopsis
133	7	1.6	91	16	Q9CB09	Q9cb09 lactococcus	206	7	1.6	211	17	Q29130	Q29130 archaeoglob
134	7	1.6	92	9	Q9MCG5	Q9mcg5 bacterioph	207	7	1.6	214	10	Q9SUA6	Q9sua6 arabidopsis
135	7	1.6	92	10	Q64820	Q64820 arabidopsis	208	7	1.6	214	16	Q8VIR3	Q8vir3 mycobacteri
136	7	1.6	99	8	Q9GDF6	Q9gdf6 sorbus torm	209	7	1.6	214	16	Q8RI83	Q8ri83 fusbacteri
137	7	1.6	101	9	Q92X17	Q92x17 mycobacteri	210	7	1.6	214	16	Q8RF02	Q8rf02 fusbacteri
138	7	1.6	102	8	Q8SLI7	Q8sl17 physocarpus	211	7	1.6	218	10	Q93WN6	Q93wn6 oryza sativ
139	7	1.6	102	8	Q8SLI6	Q8sl16 physocarpus	212	7	1.6	219	12	Q55579	Q55579 leucania se
140	7	1.6	102	8	Q8SLI5	Q8sl15 neillia aff	213	7	1.6	226	17	Q51760	Q51760 pyrococcus
141	7	1.6	102	8	Q8SLI4	Q8sl14 neillia gra	214	7	1.6	227	3	Q94348	Q94348 schizosacch
142	7	1.6	102	8	Q8SEI2	Q8sei2 neillia thy	215	7	1.6	227	4	Q9NTA4	Q9nta4 homo sapien
143	7	1.6	109	5	Q77411	Q77411 plasmodium	216	7	1.6	227	16	Q8UCQ9	Q8ucq9 agrobacteri
144	7	1.6	109	5	Q9TY36	Q9ty36 plasmodium	217	7	1.6	242	16	Q9L3M4	Q9l3m4 rhizobium l
145	7	1.6	109	10	Q9ZUJ9	Q9zuj9 arabidopsis	218	7	1.6	249	16	Q98Q03	Q98q03 mycoplasma
146	7	1.6	109	11	Q9D6B9	Q9d6b9 mus musculu	219	7	1.6	250	16	Q97DG8	Q97dg8 clostridium
147	7	1.6	112	17	Q58272	Q58272 pyrococcus	220	7	1.6	253	2	Q9EZH3	Q9ezh3 isosphaera
148	7	1.6	120	10	Q9ZUP0	Q9zup0 arabidopsis	221	7	1.6	265	10	Q9LX31	Q9lx31 arabidopsis
149	7	1.6	122	10	Q9C8V8	Q9c8v8 arabidopsis	222	7	1.6	265	10	Q8VZR4	Q8vzr4 arabidopsis
150	7	1.6	123	8	Q9G0V8	Q9g0v8 malus domes	223	7	1.6	265	10	Q8YBR6	Q8ybr6 silene roth
151	7	1.6	123	8	Q9GDF8	Q9gdf8 malus sieve	224	7	1.6	270	16	Q8RD77	Q8rd77 thermoanaer
152	7	1.6	124	8	Q9GDF4	Q9gdf4 malus micro	225	7	1.6	276	5	Q45547	Q45547 caenorhabdi
153	7	1.6	125	8	Q9GDF9	Q9gdf9 malus sieve	226	7	1.6	276	10	Q94BC4	Q94bc4 petiveria a
154	7	1.6	129	8	Q9GDF7	Q9gdf7 malus sieve	227	7	1.6	277	10	Q94BL4	Q94bl4 agdestis cl
155	7	1.6	131	8	Q9GPG3	Q9gpg3 malus niedz	228	7	1.6	277	10	Q94BA1	Q94ba1 saponaria o
156	7	1.6	133	8	Q9GDP8	Q9gdp8 malus syvie	229	7	1.6	278	16	Q50494	Q50494 streptomyce
157	7	1.6	133	8	Q9GDP7	Q9gdp7 malus sieve	230	7	1.6	279	10	Q94B86	Q94b86 trichodiade
158	7	1.6	133	8	Q9GDH2	Q9gdh2 malus asiat	231	7	1.6	280	10	Q94BB4	Q94bb4 psilotrichu
159	7	1.6	133	8	Q9GDG0	Q9gdg0 malus siebo	232	7	1.6	280	10	Q94B94	Q94b94 suaeda vera
160	7	1.6	133	8	Q9GDF5	Q9gdf5 malus torin	233	7	1.6	281	10	Q94BF5	Q94bf5 gypsophila
161	7	1.6	134	6	Q9GM51	Q9gm51 sus scrofa	234	7	1.6	281	10	Q94BC8	Q94bc8 oxybaphus n
162	7	1.6	134	8	Q9GDG9	Q9gdg9 malus fusca	235	7	1.6	281	10	Q94BC0	Q94bc0 pisonia umb

236	7	1.6	282	10	Q94BK9	Q94bk9 alternanthe	309	7	1.6	391	5	Q9VKZ1	Q9vkz1 drosophila
237	7	1.6	282	10	Q94BK4	Q94bk4 armeria mar	310	7	1.6	401	2	O54672	O54672 lactococcus
238	7	1.6	283	10	Q94BL1	Q94bl1 alluaudia a	311	7	1.6	401	2	O48702	O48702 lactococcus
239	7	1.6	283	10	Q94BD7	Q94bd7 moehringia	312	7	1.6	401	2	Q93K23	Q93k23 lactococcus
240	7	1.6	284	10	Q94B7	Q94b7 beta triglycn	313	7	1.6	405	17	Q980Y0	Q980y0 sulfolobus
241	7	1.6	284	16	Q97HD7	Q97hd7 clostridium	314	7	1.6	406	16	Q8RA81	Q8ra81 thermoanaer
242	7	1.6	285	10	Q94BE0	Q94be0 silene coro	315	7	1.6	407	5	Q8T106	Q8t106 bombyx mori
243	7	1.6	289	4	Q9Y387	Q9y387 homo sapien	316	7	1.6	409	12	Q9QA13	Q9qa13 murid herpe
244	7	1.6	291	10	Q94BJ0	Q94bj0 calandrinia	317	7	1.6	411	5	Q9SV1	Q9sv1 drosophila
245	7	1.6	293	10	Q94BL0	Q94bl0 alluaudiops	318	7	1.6	413	5	Q9N8L7	Q9n8l7 trypanosoma
246	7	1.6	294	10	Q94BD4	Q94bd4 nepenthes a	319	7	1.6	418	10	Q9FH56	Q9fh56 arabidopsis
247	7	1.6	294	2	Q93921	Q93921 synechococc	320	7	1.6	418	10	Q22995	Q22995 arabidopsis
248	7	1.6	297	5	P91314	P91314 caenorhabdi	321	7	1.6	420	3	O42980	O42980 schizosacch
249	7	1.6	302	5	Q9NEG3	Q9neg3 drosophila	322	7	1.6	423	16	Q8RIQ1	Q8riq1 fusobacteri
250	7	1.6	303	2	Q9AH90	Q9ah90 streptococc	323	7	1.6	427	3	P78807	P78807 schizosacch
251	7	1.6	311	5	O76700	O76700 caenorhabdi	324	7	1.6	428	8	Q9SET8	Q9set8 adenia digi
252	7	1.6	313	17	O8TSE9	O8tsb9 methanosarc	325	7	1.6	429	12	P88994	P88994 murid herpe
253	7	1.6	314	10	Q9SPJ7	Q9spj7 gossypium h	326	7	1.6	433	4	Q96CW1	Q96cw1 homo sapien
254	7	1.6	318	3	Q9HEU6	Q9heue6 emericeila	327	7	1.6	438	16	Q8X4U5	Q8x4u5 escherichia
255	7	1.6	324	16	Q9X6A9	Q9x6a9 yersinia pe	328	7	1.6	443	13	Q9I863	Q9i863 fugu rubrip
256	7	1.6	327	5	Q9W4V4	Q9w4v4 drosophila	329	7	1.6	447	10	Q9FLD7	Q9fld7 arabidopsis
257	7	1.6	333	2	Q9F7R1	Q9f7r1 uncultured	330	7	1.6	449	5	Q9XZN7	Q9xzn7 asterias ru
258	7	1.6	333	10	Q9S849	Q9s849 arabidopsis	331	7	1.6	459	2	O51848	O51848 pseudomonas
259	7	1.6	336	2	Q9EIS0	Q9ey50 lactobacilli	332	7	1.6	459	2	Q9R956	Q9r956 pseudomonas
260	7	1.6	343	5	O01575	O01575 caenorhabdi	333	7	1.6	459	2	Q9K4S4	Q9k4s4 pseudomonas
261	7	1.6	344	2	P74906	P74906 thermus the	334	7	1.6	459	2	O51743	O51743 pseudomonas
262	7	1.6	348	5	Q9VVM4	Q9vvw4 drosophila	335	7	1.6	459	2	P95566	P95566 pseudomonas
263	7	1.6	348	10	O49457	O49457 arabidopsis	336	7	1.6	459	10	Q9SND7	Q9snd7 arabidopsis
264	7	1.6	349	3	Q9HEB3	Q9heb3 neurospora	337	7	1.6	459	16	Q9R6R3	Q9r6r3 thermoanaer
265	7	1.6	354	5	Q9SSW8	Q9ssw8 dictyosteli	338	7	1.6	465	10	Q9AU43	Q9au45 heuchera ca
266	7	1.6	356	10	Q9ATR4	Q9atr4 oryza sativ	339	7	1.6	466	17	Q9V2D8	Q9v2d8 pyrococcus
267	7	1.6	356	13	Q98UK5	Q98uk5 brachydanio	340	7	1.6	467	13	Q9YGC6	Q9ygc6 xenopus lae
268	7	1.6	356	13	O73679	O73679 brachydanio	341	7	1.6	471	11	Q9Z0M5	Q9z0m5 mus musculu
269	7	1.6	356	16	Q9XEL9	Q9xel9 escherichia	342	7	1.6	473	10	Q9LX54	Q9lx54 arabidopsis
270	7	1.6	357	5	O16144	O16144 ascaris suu	343	7	1.6	476	2	Q9LCK4	Q9lck4 rikeneilla m
271	7	1.6	357	16	Q9R5N5	Q9r5n5 fusobacteri	344	7	1.6	479	11	Q91WP5	Q91wp5 mus musculu
272	7	1.6	358	10	Q9ATR8	Q9att8 andropogon	345	7	1.6	480	11	Q91XEL	Q91xel mus musculu
273	7	1.6	368	16	Q9PHL2	Q9phl2 campylobact	346	7	1.6	481	11	Q91WT3	Q91wt3 mus musculu
274	7	1.6	373	16	O34625	O34625 bacillus su	347	7	1.6	481	11	Q91WT1	Q91wt1 mus musculu
275	7	1.6	374	5	Q9VNH5	Q9vnh5 drosophila	348	7	1.6	482	11	Q91X92	Q91x92 mus musculu
276	7	1.6	375	8	Q9MSC4	Q9msc4 amaranthus	349	7	1.6	484	5	Q9N487	Q9n487 caenorhabdi
277	7	1.6	376	8	Q9MSG2	Q9msg2 nepenthes v	350	7	1.6	484	11	O8VEA0	O8vea0 mus musculu
278	7	1.6	376	8	Q9MSG1	Q9msg1 nepenthes t	351	7	1.6	486	6	Q95LG8	Q95lg8 macaca fasc
279	7	1.6	376	8	Q9MSG0	Q9msg0 nepenthes t	352	7	1.6	488	11	Q91WR1	Q91wr1 mus musculu
280	7	1.6	376	8	Q9MSF9	Q9msf9 nepenthes t	353	7	1.6	489	11	O8VCX4	O8vcx4 mus musculu
281	7	1.6	376	8	Q9MSF8	Q9msf8 nepenthes m	354	7	1.6	499	8	O8WJP0	O8wjp0 prunus virg
282	7	1.6	376	8	Q9MSF7	Q9msf7 nepenthes v	355	7	1.6	500	8	O8WJP2	O8wjp2 prunus laur
283	7	1.6	376	8	Q9MSF6	Q9msf6 nepenthes a	356	7	1.6	500	16	Q9RT87	Q9rt87 deinococcus
284	7	1.6	376	8	Q9MSF5	Q9msf5 nepenthes m	357	7	1.6	501	8	Q95GP4	Q95gp4 nepenthes l
285	7	1.6	376	8	Q9MSF4	Q9msf4 nepenthes k	358	7	1.6	501	8	O8WJP9	O8wjp9 neviusia al
286	7	1.6	376	8	Q9MSF3	Q9msf3 nepenthes p	359	7	1.6	502	8	O8WJN2	O8wjn2 spiraea can
287	7	1.6	376	8	Q9MSF2	Q9msf2 nepenthes d	360	7	1.6	503	8	O95GR3	O95gr3 nepenthes m
288	7	1.6	376	8	Q9MSF1	Q9msf1 ancistrocla	361	7	1.6	503	8	O95GP8	O95gp8 nepenthes m
289	7	1.6	376	8	Q9MSF0	Q9msf0 ancistrocla	362	7	1.6	503	8	Q9SE99	Q9se99 lobivia pen
290	7	1.6	376	8	Q9MSE9	Q9mse9 ancistrocla	363	7	1.6	503	8	Q9SE98	Q9se98 harrisia po
291	7	1.6	376	8	Q9MSE8	Q9mse8 ancistrocla	364	7	1.6	503	16	Q9A4M2	Q9a4m2 caulobacter
292	7	1.6	376	8	Q9MSE7	Q9mse7 triphophyll	365	7	1.6	504	6	Q9BDM1	Q9bdm1 pongo pygma
293	7	1.6	376	8	Q9MSE6	Q9mse6 dioncophyll	366	7	1.6	504	8	Q95GV7	Q95gv7 nepenthes g
294	7	1.6	376	8	Q9MSE5	Q9mse5 habroptetalu	367	7	1.6	504	8	Q95GV6	Q95gv6 nepenthes b
295	7	1.6	376	8	Q9MSD4	Q9msd4 armeria gad	368	7	1.6	504	8	Q95GV5	Q95gv5 nepenthes a
296	7	1.6	376	8	Q9MSC7	Q9msc7 simmondsia	369	7	1.6	504	8	Q95GV4	Q95gv4 nepenthes e
297	7	1.6	376	8	Q9MSC6	Q9msc6 chenopodium	370	7	1.6	504	8	Q95GV3	Q95gv3 nepenthes s
298	7	1.6	376	8	Q9MSC5	Q9msc5 bougainvill	371	7	1.6	504	8	Q95GV2	Q95gv2 nepenthes d
299	7	1.6	380	2	Q9EZH4	Q9ezh4 pirellula m	372	7	1.6	504	8	Q95GV1	Q95gv1 nepenthes i
300	7	1.6	382	17	Q97W74	Q97w74 sulfolobus	373	7	1.6	504	8	Q95GV0	Q95gv0 nepenthes l
301	7	1.6	385	2	Q48696	Q48696 lactococcus	374	7	1.6	504	8	Q95GU9	Q95gu9 nepenthes o
302	7	1.6	386	2	Q9AC99	Q9ac99 lactococcus	375	7	1.6	504	8	Q95GU8	Q95gu8 nepenthes l
303	7	1.6	388	2	O54685	O54685 lactococcus	376	7	1.6	504	8	Q95GU6	Q95gu6 nepenthes c
304	7	1.6	388	2	Q48697	Q48697 lactococcus	377	7	1.6	504	8	Q95GU5	Q95gu5 nepenthes c
305	7	1.6	388	5	Q95R32	Q95r32 drosophila	378	7	1.6	504	8	Q95GU4	Q95gu4 nepenthes r
306	7	1.6	388	5	Q9VF24	Q9vf24 drosophila	379	7	1.6	504	8	Q95GU3	Q95gu3 nepenthes r
307	7	1.6	388	10	Q941M8	Q941m8 oryza sativ	380	7	1.6	504	8	Q95GU2	Q95gu2 nepenthes i
308	7	1.6	389	10	Q9LKW6	Q9lk76 arabidopsis	381	7	1.6	504	8	Q95GU1	Q95gu1 nepenthes i

382	7	1.6	504	8	095G00	095gu0 nepenthes m	455	7	1.6	507	8	095EA4	095ea4 brownlingia
383	7	1.6	504	8	095GT9	095gt9 nepenthes m	456	7	1.6	507	8	095EA2	095ea2 coleocephal
384	7	1.6	504	8	095GT8	095gt8 nepenthes p	457	7	1.6	507	8	095EA1	095ea1 uebelmannia
385	7	1.6	504	8	095GT7	095gt7 nepenthes d	458	7	1.6	507	8	095EA0	095ea0 trichocereu
386	7	1.6	504	8	095GT6	095gt6 nepenthes k	459	7	1.6	507	8	095E97	095e97 rauhocereus
387	7	1.6	504	8	095GT5	095gt5 nepenthes a	460	7	1.6	507	8	095E96	095e96 matucana in
388	7	1.6	504	8	095GT4	095gt4 nepenthes h	461	7	1.6	507	8	095E93	095e93 parodia maa
389	7	1.6	504	8	095GT3	095gt3 nepenthes t	462	7	1.6	507	8	095E90	095e90 eriosyce au
390	7	1.6	504	8	095GT2	095gt2 nepenthes a	463	7	1.6	507	8	095E89	095e89 eriosyce is
391	7	1.6	504	8	095GT1	095gt1 nepenthes v	464	7	1.6	507	8	095E88	095e88 eriosyce su
392	7	1.6	504	8	095GT0	095gt0 nepenthes t	465	7	1.6	507	8	095E87	095e87 neowerderma
393	7	1.6	504	8	095GS9	095gs9 nepenthes m	466	7	1.6	507	8	094F94	094f94 acanthocaly
394	7	1.6	504	8	095GS8	095gs8 nepenthes v	467	7	1.6	507	8	094P94	094p94 eriosyce na
395	7	1.6	504	8	095GS7	095gs7 nepenthes n	468	7	1.6	507	8	094P31	094p31 haageocereu
396	7	1.6	504	8	095GS6	095gs6 nepenthes v	469	7	1.6	507	8	094NC4	094nc4 stetsonia c
397	7	1.6	504	8	095GS5	095gs5 nepenthes t	470	7	1.6	508	8	095EB2	095eb2 pfeiffera m
398	7	1.6	504	8	095GS4	095gs4 nepenthes t	471	7	1.6	509	8	095EB3	095eb3 pereskia gu
399	7	1.6	504	8	095GR9	095gr9 nepenthes t	472	7	1.6	509	8	095EB2	095eb2 pereskia st
400	7	1.6	504	8	095GR8	095gr8 nepenthes l	473	7	1.6	509	8	095EB1	095eb1 pereskia zi
401	7	1.6	504	8	095GR7	095gr7 nepenthes e	474	7	1.6	509	8	095EE0	095ee0 austrocylin
402	7	1.6	504	8	095GR6	095gr6 nepenthes r	475	7	1.6	509	8	095ED9	095ed9 opuntia qui
403	7	1.6	504	8	095GR5	095gr5 nepenthes p	476	7	1.6	509	8	095ED5	095ed5 blossfeldia
404	7	1.6	504	8	095GR4	095gr4 nepenthes m	477	7	1.6	509	8	095ED4	095ed4 blossfeldia
405	7	1.6	504	8	095GR1	095gr1 nepenthes d	478	7	1.6	509	8	095ED3	095ed3 frailea gra
406	7	1.6	504	8	095GR0	095gr0 nepenthes f	479	7	1.6	509	8	095ED2	095ed2 frailea pha
407	7	1.6	504	8	095GQ9	095gq9 nepenthes p	480	7	1.6	509	8	095EC7	095ec7 calymmanthi
408	7	1.6	504	8	095GQ8	095gq8 nepenthes m	481	7	1.6	509	8	095EC3	095ec3 acanthocereu
409	7	1.6	504	8	095GQ7	095gq7 nepenthes b	482	7	1.6	509	8	095EC2	095ec2 armatocereu
410	7	1.6	504	8	095GQ6	095gq6 nepenthes s	483	7	1.6	509	8	095EC1	095ec1 leptocereus
411	7	1.6	504	8	095GQ5	095gq5 nepenthes s	484	7	1.6	509	8	095EC0	095ec0 castellanos
412	7	1.6	504	8	095GQ4	095gq4 nepenthes t	485	7	1.6	509	8	095EB9	095eb9 neoraimondi
413	7	1.6	504	8	095GQ3	095gq3 nepenthes v	486	7	1.6	509	8	095EB8	095eb8 austrocactu
414	7	1.6	504	8	095GQ2	095gq2 nepenthes d	487	7	1.6	509	8	095EB7	095eb7 eulychnia l
415	7	1.6	504	8	095GQ0	095gq0 nepenthes s	488	7	1.6	509	8	095EB6	095eb6 corryocactu
416	7	1.6	504	8	095GP9	095gp9 nepenthes e	489	7	1.6	509	8	095EB5	095eb5 corryocactu
417	7	1.6	504	8	095GP5	095gp5 nepenthes m	490	7	1.6	509	8	095EB4	095eb4 pfeiffera l
418	7	1.6	504	8	095GP3	095gp3 nepenthes f	491	7	1.6	509	8	095EB3	095eb3 pfeiffera m
419	7	1.6	504	8	095GP2	095gp2 nepenthes s	492	7	1.6	509	8	095EB1	095eb1 echinocereu
420	7	1.6	504	8	094Q55	094q55 nepenthes g	493	7	1.6	509	8	095EB0	095eb0 escontria c
421	7	1.6	504	8	094PH2	094ph2 nepenthes r	494	7	1.6	509	8	095EA9	095ea9 lophocereus
422	7	1.6	504	8	094NV9	094nv9 nepenthes r	495	7	1.6	509	8	095EA8	095ea8 hylocereus
423	7	1.6	504	8	094NR5	094nr5 nepenthes b	496	7	1.6	509	8	095EA7	095ea7 selenicereu
424	7	1.6	504	8	08WJR5	08wjr5 adenostoma	497	7	1.6	509	8	095EA6	095ea6 disocactus
425	7	1.6	504	8	08WJR4	08wjr4 aruncus dio	498	7	1.6	509	8	095E92	095e92 parodia mic
426	7	1.6	504	8	08WJR1	08wjr1 chamaebatia	499	7	1.6	509	8	095E91	095e91 parodia ott
427	7	1.6	504	8	08WJR0	08wjr0 cotoneaster	500	7	1.6	509	8	095E86	095e86 hatiora sal
428	7	1.6	504	8	08WJQ8	08wjq8 exochorda r	501	7	1.6	509	8	095E85	095e85 rhipsalis f
429	7	1.6	504	8	08WJQ4	08wjq4 gillenia tr	502	7	1.6	509	8	095E84	095e84 schlumberge
430	7	1.6	504	8	08WJQ2	08wjq2 kageneckia	503	7	1.6	509	8	095E83	095e83 leptismium c
431	7	1.6	504	8	08WJQ1	08wjq1 lyonothamn	504	7	1.6	510	8	095E85	095e85 granamia br
432	7	1.6	504	8	08WJQ0	08wjq0 neillia thy	505	7	1.6	510	8	08WKL8	08wkl8 populum nig
433	7	1.6	504	8	08WJF8	08wjf8 oemleria ce	506	7	1.6	510	16	09CKF9	09ckf9 pasteurellia
434	7	1.6	504	8	08WJP7	08wjp7 photinia se	507	7	1.6	511	5	08WQO4	08wqo4 calliphora
435	7	1.6	504	8	08WJP6	08wjp6 physocarpus	508	7	1.6	511	8	095ED7	095ed7 maihuenia p
436	7	1.6	504	8	08WJP4	08wjp4 prinsepia s	509	7	1.6	511	8	095ED6	095ed6 maihuenia p
437	7	1.6	504	8	08WJN8	08wjn8 pyrus tauca	510	7	1.6	512	8	095GU7	095gu7 nepenthes l
438	7	1.6	504	8	08WJN6	08wjn6 rhodotypos	511	7	1.6	514	8	095GP1	095gp1 ancistrocla
439	7	1.6	504	8	08WJN4	08wjn4 sorbaria so	512	7	1.6	517	5	09VX85	09vx85 drosophila
440	7	1.6	504	8	08WJN3	08wjn3 sorbus cali	513	7	1.6	520	16	08XKT0	08xkt0 clostridium
441	7	1.6	504	8	08WJN0	08wjn0 vaquelinia	514	7	1.6	523	11	09QXU8	09qxu8 rattus norv
442	7	1.6	504	16	092LK8	092lk8 rhizobium m	515	7	1.6	524	4	09H188	09h188 homo sapien
443	7	1.6	505	8	095EA3	095ea3 gymnocalyci	516	7	1.6	525	4	09H187	09h187 homo sapien
444	7	1.6	505	8	095E95	095e95 parodia has	517	7	1.6	525	4	09H186	09h186 homo sapien
445	7	1.6	505	8	095E94	095e94 parodia bue	518	7	1.6	525	4	09H185	09h185 homo sapien
446	7	1.6	506	8	08WJQ9	08wjq9 crataegus m	519	7	1.6	525	4	09H184	09h184 homo sapien
447	7	1.6	506	8	08WJQ3	08wjq3 holodiscus	520	7	1.6	525	4	09H183	09h183 homo sapien
448	7	1.6	506	8	08WJP3	08wjp3 prunus dulc	521	7	1.6	525	4	09H182	09h182 homo sapien
449	7	1.6	506	8	08WJP1	08wjp1 prunus pers	522	7	1.6	525	4	09H181	09h181 homo sapien
450	7	1.6	507	8	095ED8	095ed8 peresklopsi	523	7	1.6	525	6	09BG53	09bg53 pan trogloc
451	7	1.6	507	8	095EC6	095ec6 copiapoa so	524	7	1.6	528	16	09RTP5	09rtp5 deinococcus
452	7	1.6	507	8	095EC5	095ec5 copiapoa br	525	7	1.6	529	5	095Z26	095z26 caenorhabdi
453	7	1.6	507	8	095EC4	095ec4 copiapoa la	526	7	1.6	531	3	04Z797	04z797 podospora a
454	7	1.6	507	8	095EA5	095ea5 brownlingia	527	7	1.6	532	5	017500	017500 caenorhabdi

528	7	1.6	538	10	Q9XFM4	Q9xfm4 fagopyrum e	601	7	1.6	905	13	Q90505	Q90505 fundulus he
529	7	1.6	539	10	Q94LS6	Q94ls6 oryza sativ	602	7	1.6	907	16	Q8y0t5	Q8y0t5 ralstonia s
530	7	1.6	546	10	Q08737	Q08737 triticum ae	603	7	1.6	913	16	Q9PQ03	Q9pq03 ureaplasma
531	7	1.6	550	3	Q9P5P9	Q9p5p9 neurospora	604	7	1.6	929	5	Q9NGW5	Q9ngw5 drosophila
532	7	1.6	553	3	Q9WQP6	Q9wqp6 drosophila	605	7	1.6	929	5	Q9NBL3	Q9nbl3 drosophila
533	7	1.6	554	2	Q9ZFF1	Q9zff1 klebsiella	606	7	1.6	929	5	Q9NBL3	Q9nbl3 drosophila
534	7	1.6	554	2	Q59470	Q59470 klebsiella	607	7	1.6	938	10	Q9LDD7	Q9ldd7 lycopersico
535	7	1.6	554	16	Q8Z5M2	Q8z5m2 salmonella	608	7	1.6	939	5	Q9NHQ0	Q9nhq0 drosophila
536	7	1.6	555	5	Q16745	Q16745 caenorhabdi	609	7	1.6	950	5	Q95TRA	Q95tr4 drosophila
537	7	1.6	573	16	Q926K8	Q926k8 listeria in	610	7	1.6	951	10	Q9AR52	Q9ar52 vicia faba
538	7	1.6	579	2	Q93LM7	Q93lm7 pseudomonas	611	7	1.6	951	10	Q43106	Q43106 phaseolus v
539	7	1.6	587	10	Q9ZWF9	Q9zwf9 cucumis mel	612	7	1.6	956	10	Q43131	Q43131 vicia faba
540	7	1.6	588	5	Q9NCM4	Q9ncm4 plasmodium	613	7	1.6	956	10	Q42932	Q42932 nicotiana p
541	7	1.6	588	5	Q9NCM2	Q9ncm2 plasmodium	614	7	1.6	956	10	Q43182	Q43182 solanum tub
542	7	1.6	595	5	Q9NCM1	Q9ncm1 plasmodium	615	7	1.6	956	10	Q9SPD5	Q9spd5 lycopersico
543	7	1.6	595	5	Q9NCM0	Q9ncm0 plasmodium	616	7	1.6	957	10	Q43002	Q43002 oryza sativ
544	7	1.6	596	5	Q19953	Q19953 caenorhabdi	617	7	1.6	958	4	Q96A52	Q96a52 homo sapien
545	7	1.6	597	5	Q9N657	Q9n657 plasmodium	618	7	1.6	974	5	Q20143	Q20143 caenorhabdi
546	7	1.6	597	5	Q9NCN3	Q9ncn3 plasmodium	619	7	1.6	989	10	Q9SCM5	Q9scm5 arabidopsis
547	7	1.6	597	5	Q9NCN1	Q9ncn1 plasmodium	620	7	1.6	1014	12	Q8Q556	Q8qs56 chimpanzee
548	7	1.6	597	5	Q9NCM6	Q9ncm6 plasmodium	621	7	1.6	1019	5	Q76143	Q76143 drosophila
549	7	1.6	597	5	Q9NCM3	Q9ncm3 plasmodium	622	7	1.6	1021	5	Q77289	Q77289 drosophila
550	7	1.6	599	5	Q9NCN2	Q9ncn2 plasmodium	623	7	1.6	1024	5	Q9W268	Q9w268 drosophila
551	7	1.6	599	5	Q9NCM7	Q9ncm7 plasmodium	624	7	1.6	1054	11	Q9JM99	Q9jlm99 mus musculu
552	7	1.6	599	5	Q9NCM5	Q9ncm5 plasmodium	625	7	1.6	1063	3	Q96WN2	Q96wn2 paracoccidi
553	7	1.6	609	13	Q9W6M5	Q9w6m5 agkistrodon	626	7	1.6	1083	5	Q8T3K8	Q8t3k8 drosophila
554	7	1.6	610	12	Q71123	Q71123 rhesus cyto	627	7	1.6	1104	16	Q51934	Q51934 thermotoga
555	7	1.6	613	5	Q9NCN0	Q9ncn0 plasmodium	628	7	1.6	1122	5	Q23682	Q23682 caenorhabdi
556	7	1.6	614	5	Q9NCM9	Q9ncm9 plasmodium	629	7	1.6	1161	5	Q9W2X8	Q9w2x8 drosophila
557	7	1.6	616	5	Q9NCM8	Q9ncm8 plasmodium	630	7	1.6	1173	12	Q98177	Q98177 molluscum c
558	7	1.6	624	16	Q99Q28	Q99q28 staphylococ	631	7	1.6	1175	12	Q85285	Q85285 molluscum c
559	7	1.6	630	2	Q9EVD6	Q9evd6 actinomyces	632	7	1.6	1180	5	Q24163	Q24163 drosophila
560	7	1.6	638	5	Q9VYX9	Q9vyx9 drosophila	633	7	1.6	1186	5	Q9VED3	Q9ved3 drosophila
561	7	1.6	642	16	Q9JYK7	Q9jyk7 neisseria m	634	7	1.6	1208	5	Q9U4W1	Q9u4w1 aedes aegyp
562	7	1.6	642	16	Q9JTK6	Q9jtk6 neisseria m	635	7	1.6	1214	10	Q9LD60	Q9ld60 arabidopsis
563	7	1.6	659	5	Q17792	Q17792 caenorhabdi	636	7	1.6	1263	4	Q15044	Q15044 homo sapien
564	7	1.6	666	2	Q32495	Q32495 bacteroides	637	7	1.6	1283	5	Q95RH4	Q95rh4 drosophila
565	7	1.6	666	5	Q9VP48	Q9vp48 drosophila	638	7	1.6	1284	5	Q9VJ79	Q9vj79 drosophila
566	7	1.6	670	5	Q9N554	Q9n554 caenorhabdi	639	7	1.6	1308	5	Q01924	Q01924 caenorhabdi
567	7	1.6	684	2	Q65990	Q65990 clostridium	640	7	1.6	1339	5	Q26048	Q26048 pacifastacu
568	7	1.6	684	16	Q97MN8	Q97mn8 clostridium	641	7	1.6	1441	5	Q96957	Q96957 drosophila
569	7	1.6	685	4	Q96CV7	Q96cv7 homo sapien	642	7	1.6	1446	5	Q77063	Q77063 aplysia cal
570	7	1.6	688	10	Q9S8Y0	Q9s8y0 atropa bell	643	7	1.6	1461	5	Q8SQ03	Q8sq03 encephalito
571	7	1.6	700	2	Q34003	Q34003 rhodobacter	644	7	1.6	1465	5	Q9VMD0	Q9vmd0 drosophila
572	7	1.6	706	5	Q9V417	Q9v417 drosophila	645	7	1.6	1521	5	Q8SX83	Q8sx83 drosophila
573	7	1.6	706	5	Q8SXG1	Q8sxg1 drosophila	646	7	1.6	1534	5	Q9W3D3	Q9w3d3 drosophila
574	7	1.6	718	5	Q9VRH4	Q9vrh4 drosophila	647	7	1.6	1561	5	Q9W3D2	Q9w3d2 drosophila
575	7	1.6	737	16	Q8R5Z1	Q8r5z1 fusobacteri	648	7	1.6	1591	3	Q9HFW1	Q9hfw1 ashbya goss
576	7	1.6	758	5	Q96037	Q96037 clona savig	649	7	1.6	1597	5	Q61346	Q61346 drosophila
577	7	1.6	768	16	Q9YANO	Q9yan0 listeria mo	650	7	1.6	1625	10	Q08367	Q08367 zea mays (m
578	7	1.6	784	6	Q9TUN5	Q9tun5 sus scrofa	651	7	1.6	1645	5	Q8WSS9	Q8wss9 drosophila
579	7	1.6	784	6	Q95JH1	Q95jh1 sus scrofa	652	7	1.6	1677	5	Q9BKV5	Q9bkv5 leishmania
580	7	1.6	792	16	Q9PE24	Q9pez4 xylella fas	653	7	1.6	1685	10	Q43248	Q43248 zea mays (m
581	7	1.6	799	17	Q8TH11	Q8th11 pyrococcus	654	7	1.6	1726	5	Q02569	Q02569 plasmodium
582	7	1.6	816	5	Q26026	Q26026 plasmodium	655	7	1.6	1751	5	Q26194	Q26194 plasmodium
583	7	1.6	818	13	Q9PSV9	Q9psv9 xenopus lae	656	7	1.6	1895	5	Q9VMS5	Q9vms5 drosophila
584	7	1.6	818	13	Q91742	Q91742 xenopus lae	657	7	1.6	1911	5	Q9W4M7	Q9w4m7 drosophila
585	7	1.6	828	13	Q9DQK3	Q9dqk3 xenopus lae	658	7	1.6	2038	5	Q967Y0	Q967y0 dictyosteli
586	7	1.6	828	13	Q91743	Q91743 xenopus lae	659	7	1.6	2091	5	Q9VJ69	Q9vj69 drosophila
587	7	1.6	829	13	Q9PSV8	Q9psv8 xenopus lae	660	7	1.6	2280	5	Q9V8E6	Q9v8e6 drosophila
588	7	1.6	832	15	Q93258	Q93258 chimpanzee	661	7	1.6	2302	5	Q9N693	Q9n693 drosophila
589	7	1.6	842	2	Q53112	Q53112 mycoplasma	662	7	1.6	2310	5	Q9GRA9	Q9gra9 drosophila
590	7	1.6	846	2	Q9ZFP7	Q9zfp7 vibrio chol	663	7	1.6	2311	10	Q48959	Q48959 triticum ae
591	7	1.6	846	16	Q9KQP6	Q9kqp6 vibrio chol	664	7	1.6	2321	10	Q947M6	Q947m6 setaria ita
592	7	1.6	851	13	Q42507	Q42507 xenopus lae	665	7	1.6	2325	10	Q41743	Q41743 zea mays (m
593	7	1.6	856	5	Q8SSV0	Q8ssv0 dictyosteli	666	7	1.6	3080	5	Q9VRY3	Q9vry3 drosophila
594	7	1.6	861	5	Q9VRT8	Q9vrt8 drosophila	667	7	1.6	5476	5	Q9NJ17	Q9nj17 drosophila
595	7	1.6	865	16	Q8XY19	Q8xy19 ralstonia s	668	7	1.6	5533	5	Q9VPL2	Q9vpl2 drosophila
596	7	1.6	869	4	Q15572	Q15572 homo sapien	669	7	1.6	5533	5	Q9U6C3	Q9u6c3 drosophila
597	7	1.6	876	10	Q9FZB0	Q9fzb0 arabidopsis	670	7	1.6	5554	5	Q9NHN1	Q9nhn1 drosophila
598	7	1.6	885	16	Q92UL6	Q92ul6 rhizobium m	671	7	1.6	5560	5	Q9VPL1	Q9vpl1 drosophila
599	7	1.6	885	16	Q8XTK5	Q8xtk5 ralstonia s	672	7	1.6	5616	2	Q48439	Q48439 klebsiella
600	7	1.6	897	16	Q98Q69	Q98q69 mycoplasma	673	7	1.4	16	16	Q8X8T5	Q8x8t5 escherichia

674	6	1.4	29	4	Q9UM98	Q9um98 homo sapien	747	6	1.4	89	13	093572	093572 ambystoma m
675	6	1.4	29	6	Q9TRL6	Q9trl6 bos taurus	748	6	1.4	92	5	Q18410	Q18410 caenorhabdi
676	6	1.4	32	16	Q9KKK1	Q9kkk1 vibrio chol	749	6	1.4	93	10	Q9SL66	Q9sl66 arabidopsis
677	6	1.4	38	4	Q16097	Q16097 homo sapien	750	6	1.4	93	10	Q9L2N9	Q9lzn9 arabidopsis
678	6	1.4	39	4	Q16098	Q16098 homo sapien	751	6	1.4	93	16	Q926J5	Q929j5 rickettsia
679	6	1.4	46	17	Q8U113	Q8u113 pyrococcus	752	6	1.4	94	2	Q9SL39	Q9sl39 escherichia
680	6	1.4	48	10	Q8SAS1	Q8sas1 pinus sylve	753	6	1.4	94	4	Q14814	Q14814 homo sapien
681	6	1.4	48	10	Q8SAS0	Q8sas0 pinus sylve	754	6	1.4	94	5	Q9U709	Q9u709 plasmodium
682	6	1.4	50	12	Q996F7	Q996f7 human echov	755	6	1.4	94	5	Q9U708	Q9u708 plasmodium
683	6	1.4	51	4	Q96BA5	Q96ba5 homo sapien	756	6	1.4	94	5	Q9U706	Q9u706 plasmodium
684	6	1.4	51	5	Q268A5	Q268a5 trypanosoma	757	6	1.4	94	5	Q9U705	Q9u705 plasmodium
685	6	1.4	51	5	Q9W5B2	Q9w5b2 drosophila	758	6	1.4	95	5	Q9U712	Q9u712 plasmodium
686	6	1.4	52	17	Q97ZP0	Q97zp0 sulfolobus	759	6	1.4	95	5	Q9U711	Q9u711 plasmodium
687	6	1.4	54	4	P78495	P78495 homo sapien	760	6	1.4	95	5	Q9U710	Q9u710 plasmodium
688	6	1.4	54	16	Q8X3B5	Q8x3b5 escherichia	761	6	1.4	95	5	Q9U707	Q9u707 plasmodium
689	6	1.4	55	17	Q8ZWU5	Q8zwu5 pyrobaculum	762	6	1.4	95	5	Q9U704	Q9u704 plasmodium
690	6	1.4	56	5	Q8T2R1	Q8t2r1 dictyosteli	763	6	1.4	95	15	Q88340	Q88340 simian t-ly
691	6	1.4	56	16	Q8YVT1	Q8yvt1 anabaena sp	764	6	1.4	95	15	Q88342	Q88342 simian t-ly
692	6	1.4	57	16	Q9PNN4	Q9pnn4 campylobact	765	6	1.4	96	2	O33423	O33423 pseudomonas
693	6	1.4	59	2	Q9RQF7	Q9rqf7 vibrio chol	766	6	1.4	96	5	O61170	O61170 plasmodium
694	6	1.4	59	9	Q9MBU8	Q9mbu8 vibrio phag	767	6	1.4	97	13	Q9W730	Q9w730 acipenser r
695	6	1.4	60	5	Q9VSB4	Q9vsb4 drosophila	768	6	1.4	97	17	Q976D2	Q976d2 sulfolobus
696	6	1.4	60	16	Q8Z523	Q8zs23 anabaena sp	769	6	1.4	99	2	O07070	O07070 burkholderi
697	6	1.4	60	16	Q8VK73	Q8vk73 mycobacteri	770	6	1.4	99	2	Q8RP91	Q8rp91 cylindrospe
698	6	1.4	60	17	Q9HMB5	Q9hmb5 halobacteri	771	6	1.4	99	17	Q97Z39	Q97z39 sulfolobus
699	6	1.4	61	12	Q8VAM4	Q8vaw4 white spot	772	6	1.4	100	10	Q8RUF0	Q8ruf0 oryza sativ
700	6	1.4	63	5	Q9VDX0	Q9vdx0 drosophila	773	6	1.4	101	16	Q9FBZ6	Q9fbz6 streptomyce
701	6	1.4	63	16	Q8R7R7	Q8r7r7 thermoaenar	774	6	1.4	102	5	Q94189	Q94189 caenorhabdi
702	6	1.4	64	10	Q40049	Q40049 hordeum vul	775	6	1.4	102	5	Q9VUE1	Q9vuel drosophila
703	6	1.4	64	13	Q91220	Q91220 oncorhynch	776	6	1.4	102	5	Q9VED9	Q9ved9 drosophila
704	6	1.4	67	13	Q92165	Q92165 oncorhynch	777	6	1.4	102	8	Q8SDZ2	Q8sdz2 neillia thi
705	6	1.4	67	13	Q91235	Q91235 oncorhynch	778	6	1.4	102	8	Q8SDZ1	Q8sdz1 neillia thi
706	6	1.4	67	13	Q91172	Q91172 oncorhynch	779	6	1.4	102	8	Q8SD20	Q8sdz0 neillia uek
707	6	1.4	67	13	Q91158	Q91158 oncorhynch	780	6	1.4	102	8	Q8SDY9	Q8sdy9 stephanandr
708	6	1.4	67	13	Q92166	Q92166 oncorhynch	781	6	1.4	102	8	Q8SDY8	Q8sdy8 stephanandr
709	6	1.4	69	10	Q8S2R7	Q8s2r7 thellungiel	782	6	1.4	102	8	Q8SDY7	Q8sdy7 stephanandr
710	6	1.4	69	13	Q9P765	Q9pt65 xenopus lae	783	6	1.4	102	10	Q9LVC7	Q9lvc7 arabidopsis
711	6	1.4	69	13	Q13044	Q13044 scyllorhinu	784	6	1.4	102	11	P97294	P97294 mus musculu
712	6	1.4	70	5	Q9V374	Q9v37 drosophila	785	6	1.4	102	16	Q92J02	Q92j02 rickettsia
713	6	1.4	70	16	Q97ME7	Q97me7 clostridium	786	6	1.4	103	15	Q91084	Q91084 human immun
714	6	1.4	71	5	Q9W006	Q9w0m6 drosophila	787	6	1.4	103	17	Q970X8	Q970x8 sulfolobus
715	6	1.4	71	17	Q8TKS8	Q8tk8 methanosarc	788	6	1.4	106	2	Q9KIT8	Q9kit8 bacillus me
716	6	1.4	72	5	Q8T1A1	Q8t1a1 dictyosteli	789	6	1.4	106	2	Q9KIT4	Q9kit4 clostridium
717	6	1.4	72	16	Q25973	Q25973 helicobacte	790	6	1.4	106	10	Q9LS87	Q9ls87 arabidopsis
718	6	1.4	73	5	Q8T844	Q8t844 dictyosteli	791	6	1.4	107	5	Q9VQW5	Q9vqw5 drosophila
719	6	1.4	73	12	Q9WHW7	Q9whw7 tt virus. h	792	6	1.4	108	2	O82983	O82983 bacillus sp
720	6	1.4	75	5	O8T3D9	Q8t3d9 caenorhabdi	793	6	1.4	108	2	Q9XD75	Q9xd75 nitrobacter
721	6	1.4	75	13	Q91456	Q91456 salvelinus	794	6	1.4	108	2	O8VSW4	Q8vsw4 staphylococ
722	6	1.4	76	5	Q9VLH4	Q9vlh4 drosophila	795	6	1.4	108	16	O8XXT6	O8xxt6 ralstonia s
723	6	1.4	76	16	Q92FX4	Q92fx4 rickettsia	796	6	1.4	109	5	Q9W346	Q9w346 drosophila
724	6	1.4	76	17	Q8U1J9	Q8u1j9 pyrococcus	797	6	1.4	110	12	Q84468	Q84468 paramecium
725	6	1.4	77	5	Q20690	Q20690 caenorhabdi	798	6	1.4	110	17	O58026	O58026 pyrococcus
726	6	1.4	77	16	Q97HT2	Q97ht2 clostridium	799	6	1.4	111	5	O21111	O21111 caenorhabdi
727	6	1.4	77	16	Q9ZJ18	Q9zj18 helicobacte	800	6	1.4	113	16	Q9XJ11	Q9xj11 neisseria m
728	6	1.4	78	5	P90569	P90569 plasmodium	801	6	1.4	113	16	Q9JX91	Q9jx91 neisseria m
729	6	1.4	78	16	Q92DV6	Q92dv6 listeria in	802	6	1.4	114	5	Q18724	Q18724 caenorhabdi
730	6	1.4	78	16	O8Y943	O8y943 listeria mo	803	6	1.4	115	11	O08932	O08932 mus musculu
731	6	1.4	78	17	Q8TH48	Q8th48 methanopyru	804	6	1.4	115	17	Q9YAJ2	Q9yaj2 aeropyrum p
732	6	1.4	79	4	Q96P11	Q96p11 homo sapien	805	6	1.4	116	5	Q95242	Q95242 plasmodium
733	6	1.4	79	15	Q80853	Q80853 human t-lym	806	6	1.4	116	6	Q97933	Q97933 phocenoide
734	6	1.4	79	15	Q80854	Q80854 human t-lym	807	6	1.4	116	6	Q97934	Q97934 pseudorca c
735	6	1.4	79	15	Q80855	Q80855 human t-lym	808	6	1.4	116	11	Q8VDP99	Q8vdp99 rattus ratt
736	6	1.4	80	17	O26182	O26182 methanobact	809	6	1.4	116	16	Q9K8E3	Q9k8e3 bacillus ha
737	6	1.4	82	10	O80460	O80460 arabidopsis	810	6	1.4	117	11	Q9D1B5	Q9dlb5 mus musculu
738	6	1.4	83	12	Q89074	Q89074 variola vir	811	6	1.4	118	2	O87097	O87097 bacillus su
739	6	1.4	84	5	Q917M5	Q917m5 drosophila	812	6	1.4	118	5	Q9U4E9	Q9u4e9 aedes aegyp
740	6	1.4	84	10	Q9XFEU2	Q9xfeu2 chlamydomon	813	6	1.4	119	5	Q9NUJ9	Q9nuj9 aedes aegyp
741	6	1.4	84	16	Q97G10	Q97g10 clostridium	814	6	1.4	119	8	Q9ME92	Q9me92 oenothera h
742	6	1.4	86	5	Q9XW64	Q9xw64 caenorhabdi	815	6	1.4	119	16	Q92CB5	Q92cb5 listeria in
743	6	1.4	87	2	Q9APV1	Q9apv1 pseudomonas	816	6	1.4	119	16	Q926Z1	Q926z1 chlamydia p
744	6	1.4	87	5	Q91810	Q91810 drosophila	817	6	1.4	121	9	O8W656	O8w656 bacterioph
745	6	1.4	87	16	Q914Y3	Q914y3 pseudomonas	818	6	1.4	121	17	Q8ZV63	Q8zv63 pyrobaculum
746	6	1.4	88	5	Q9V5H9	Q9v5h9 drosophila	819	6	1.4	122	5	Q25669	Q25669 plasmodium





```

966 1.4 159 10 Q9XG40 Q9xg40 guillardia
967 1.4 159 12 Q9YMT6 Q9ymt6 lymantria d
968 1.4 160 2 Q9KI53 Q9ki53 enterococcu
969 1.4 161 3 Q12155 Q12155 saccharomyc
970 1.4 161 4 Q9NWT0 Q9nwt0 homo sapien
971 1.4 161 4 Q8WZ09 Q8wz09 homo sapien
972 1.4 161 16 Q98BL0 Q98bl0 rhizobium l
973 1.4 161 16 Q8R987 Q8r987 thermoanaer
974 1.4 162 16 Q9KDR7 Q9kdr7 bacillus ha
975 1.4 162 16 Q97Q11 Q97q11 streptococc
976 1.4 163 2 Q9ZAR6 Q9zar6 pseudomonas
977 1.4 163 8 Q9BDC2 Q9bdc2 antillocapra
978 1.4 163 8 Q9T9N9 Q9t9n9 terebratul
979 1.4 163 16 Q9JVA7 Q9jva7 neisseria m
980 1.4 163 17 Q58287 Q58287 pyrococcus
981 1.4 165 5 Q02457 Q02457 teladorsagi
982 1.4 165 8 Q35961 Q35961 oryza sativ
983 1.4 165 10 Q9SP16 Q9spi6 picea ruben
984 1.4 165 13 P87487 P87487 oncorhynchu
985 1.4 165 16 Q9PIB5 Q9piB5 campylobact
986 1.4 165 16 Q9A091 Q9a091 streptococc
987 1.4 166 1 Q93673 Q93673 sulfolobus
988 1.4 166 3 Q60166 Q60166 schizosacch
989 1.4 166 5 Q9VK17 Q9vk17 drosophila
990 1.4 166 16 Q08797 Q08797 vibrio chol
991 1.4 167 4 Q96DC5 Q96dc5 homo sapien
992 1.4 167 8 Q9MNX1 Q9mnx1 pagurus lon
993 1.4 167 10 Q9SAS3 Q9sas3 triticum ae
994 1.4 167 16 Q67111 Q67111 aquifex ae
995 1.4 167 16 Q825U9 Q825u9 salmonella
996 1.4 188 2 Q9EUC3 Q9euc3 streptococc
997 1.4 188 2 Q85258 Q85258 streptococc
998 1.4 168 2 Q85260 Q85260 streptococc
999 1.4 168 2 Q85261 Q85261 streptococc
1000 1.4 168 2 Q85262 Q85262 streptococc

```

## ALIGNMENTS

```

RESULT 1
Q25922 PRELIMINARY; PRT; 1720 AA.
AC Q25922:
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DE Precursor of the major merozoite surface antigens.
DE Plasmodium falciparum (isolate NF54).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5843;
RN [1]
RP SEQUENCE OF 1069-1720 FROM N.A.
RC STRAIN=NF54;
RA Tolle R., Bujard H., Cooper J.A.;
RL Exp. Parasitol. 0:0-0(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NF54;
RA Tolle R.;
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=NF54;
RX MEDLINE=96123395; PubMed=8577332;
RA Pan W., Tolle R., Bujard H.;
RT "A direct and rapid sequencing strategy for the Plasmodium falciparum antigen gene gp190/MSA1."
RL Mol. Biochem. Parasitol. 73:241-244(1995).
DR EMBL; Z35327; CAA84556.1; -
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW Merozoite; Signal.

```

```

FT SIGNAL 1 19 POTENTIAL.
SQ SEQUENCE 1720 AA; 195725 MW; 717B2FD1E637A8A3 CRC64;
Query Match 86.8%; Score 374; DB 5; Length 1720;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 374; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 58 AISVTMDNLSLGFENEYDVYILKPLAGYRSLKQIEKNFTFNLNDILNSRLKRRKY 117
DB 1327 AISVTMDNLSLGFENEYDVYILKPLAGYRSLKQIEKNFTFNLNDILNSRLKRRKY 1386
QY 118 FLDVLESQKQFHRHISSEYIIEDSFKLNSQKNTLLSKYIKESVENDIKRPAQEGIS 177
DB 1387 FLDVLESQKQFHRHISSEYIIEDSFKLNSQKNTLLSKYIKESVENDIKRPAQEGIS 1446
QY 178 YYEKVLAKYKDDLESIKKVIKEEKEKFPSPPTPPSPAKTDEQKESKFLPFLTNIELT 237
DB 1447 YYEKVLAKYKDDLESIKKVIKEEKEKFPSPPTPPSPAKTDEQKESKFLPFLTNIELT 1506
QY 238 YNVLNKKIDDYLYNLKAKINDCNVKEKDEAHVKITKLSDLKAIDDKIDLFKNPYDFAIKK 297
DB 1507 YNVLNKKIDDYLYNLKAKINDCNVKEKDEAHVKITKLSDLKAIDDKIDLFKNPYDFAIKK 1566
QY 298 LINDDTKKDMLGKLLTGLVQNPNTIISKLEKGFQDMLNISQHCVKKOCPPNSGCFR 357
DB 1567 LINDDTKKDMLGKLLTGLVQNPNTIISKLEKGFQDMLNISQHCVKKOCPPNSGCFR 1626
QY 358 HLDERECKLLNKKQEGDKCVENPNPTCNENNGCCDADATCTEEDSGSRKKKTCCECTK 417
DB 1627 HLDERECKLLNKKQEGDKCVENPNPTCNENNGCCDADATCTEEDSGSRKKKTCCECTK 1686
QY 418 PDSYPLFDGIFCSS 431
DB 1687 PDSYPLFDGIFCSS 1700
RESULT 2
Q25923 PRELIMINARY; PRT; 652 AA.
AC Q25923:
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DE Merozoite surface antigen 1 (Fragment).
DE Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE OF 1-298 FROM N.A.
RC STRAIN=FCH5/NF7;
RX MEDLINE=88142999; PubMed=2449612;
RA Peterson M.G., Coppel R.L., McIntyre P., Langford C.J., Woodrow G., Brown G.V., Anders R.F., Kemp D.J.;
RT "Variation in the precursor to the major merozoite surface antigens of Plasmodium falciparum."
RL Mol. Biochem. Parasitol. 27:291-302(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FCH5/NF7;
RA Tolle R.;
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=FCH5/NF7;
RX MEDLINE=95354793; PubMed=7628566;
RA Tolle R., Bujard H., Cooper J.A.;
RT "Plasmodium falciparum: variations within the C-terminal region of merozoite surface antigen-1."
RL Exp. Parasitol. 81:47-54(1995).
DR EMBL; Z35328; CAA84557.1; -
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW NON_TER 1 1

```

Qy	238	YNNLVNKIDDDYLINLAKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFRNPYDEAIKK	293
Db	181	YNNLVNKIDDDYLINLAKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFRNPYDEAIKK	293
Qy	298	LINDDTKDMGLKLLSTGLVONFPNTIISKLLIEGKFQDMLNISQHCYKKQCP	350
Db	241	LINDDTKDMGLKLLSTGLVONFPNTIISKLLIEGKFQDMLNISQHCYKKQCP	293
RESULT 4			
ID	043995	PRELIMINARY;	PRT; 373 AA.
AC	043995;		
DT	01-JUN-1998	(TREMBlrel. 06, Created)	
DT	01-JUN-1998	(TREMBlrel. 06, Last sequence update)	
DT	01-JUN-2002	(TREMBlrel. 21, Last annotation update)	
DE		Merozoite surface protein-1 (Fragment).	
GN	MSP-1.		
OS	Plasmodium falciparum.		
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
OX	NCBI_TaxID=5833;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=KENYA-1;		
RX	MEDLINE=98319411; PubMed=9657329;		
RA	Qari S.H., Shi Y.P., Goldman I.F., Nahlen B.L., Tibayrenc M.,		
RA	Lal A.A.;		
RT	"Predicted and observed alleles of Plasmodium falciparum merozoite		
RT	surface protein-1 (MSP-1), a potential malaria vaccine antigen.";		
RL	Mol. Biochem. Parasitol. 92:241-252(1998).		
DR	EMBL; AF040567; AAC39097.1; -		
DR	InterPro; IPR000561; EGF-like.		
DR	Pfam; PF00008; EGF; 1.		
FT	NON_TER	1 373	
FT	NON_TER	373 373	
SQ	SEQUENCE	373 AA; 42873 MW; 886CF169A7AF5194 CRC64;	
Query Match 53.6%; Score 231; DB 5; Length 373;			
Best Local Similarity 100.0%; Pred. No. 3.1e-232;			
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps			
Qy	58	AISVTMDNILSGFENEYDVITYLPLAGVYRSLKQIEKNIFTNLNLDILNSRLKKRY	111
Db	1	AISVTMDNILSGFENEYDVITYLPLAGVYRSLKQIEKNIFTNLNLDILNSRLKKRY	60
Qy	118	FLDVLESDLMPFKHISSEYIIESFLLNSEQNTLLKSKYIKESVENDIKFAQEGIS	178
Db	61	FLDVLESDLMPFKHISSEYIIESFLLNSEQNTLLKSKYIKESVENDIKFAQEGIS	120
Qy	178	YIEKVLAKYKDDLESIKKVIKEEKEKFPSPPTTPPSPAKTDQEKESKFLPFLTNIETL	238
Db	121	YIEKVLAKYKDDLESIKKVIKEEKEKFPSPPTTPPSPAKTDQEKESKFLPFLTNIETL	180
Qy	238	YNNLVNKIDDDYLINLAKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFRN	288
Db	181	YNNLVNKIDDDYLINLAKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFRN	231
RESULT 5			
ID	043996	PRELIMINARY;	PRT; 373 AA.
AC	043996;		
DT	01-JUN-1998	(TREMBlrel. 06, Created)	
DT	01-JUN-1998	(TREMBlrel. 06, Last sequence update)	
DT	01-DEC-2001	(TREMBlrel. 19, Last annotation update)	
DE		Merozoite surface protein-1 (Fragment).	
GN	MSP-1.		
OS	Plasmodium falciparum.		
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
OX	NCBI_TaxID=5833;		
RN	[1]		
RC	SEQUENCE FROM N.A.		
RC	STRAIN=KENYA-2;		



```

RX MEDLINE=88011243; PubMed=3079521;
RA Tanabe K., Mackay M., Goman M., Scaife J.G.;
RT "Allelic Dimorphism in a Surface Antigen Gene of the Malaria Parasite
RT Plasmodium falciparum.";
RL J. Mol. Biol. 195:273-287(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13343; BAA02604.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 570 AA; 64630 MW; 8674DEC89B2D562A CRC64;

Query Match 44.1%; Score 190; DB 5; Length 570;
Best Local Similarity 100.0%; Pred. No. 2.9e-189;
Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 TFNLNLNDILNSRLKRRKYFLDVLESFLMQFKHISSEYIIEDSFKLLNSEQKNTLLKSY 158
Db 218 TFNLNLNDILNSRLKRRKYFLDVLESFLMQFKHISSEYIIEDSFKLLNSEQKNTLLKSY 277

QY 159 KYIKESVENDIRFAQEGISYYEKVLAKYKDDLESIKKVIKEEKFPSPPTPPSPAKT 218
Db 278 KYIKESVENDIRFAQEGISYYEKVLAKYKDDLESIKKVIKEEKFPSPPTPPSPAKT 337

QY 219 DEQKESKFLPFLTNTETLYNNLVNKIDDDYLNLKAKINDCNVKEDEAHVKITKLSDLKA 278
Db 338 DEQKESKFLPFLTNTETLYNNLVNKIDDDYLNLKAKINDCNVKEDEAHVKITKLSDLKA 397

QY 279 IDDKIDLFRN 288
Db 398 IDDKIDLFRN 407

RESULT 9
Q25968 ID Q25968 PRELIMINARY; PRT; 570 AA.
AC Q25968;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13343; BAA02606.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 570 AA; 64632 MW; 424BF553CCC2F2BE CRC64;

Query Match 44.1%; Score 190; DB 5; Length 570;
Best Local Similarity 100.0%; Pred. No. 2.9e-189;
Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 TFNLNLNDILNSRLKRRKYFLDVLESFLMQFKHISSEYIIEDSFKLLNSEQKNTLLKSY 158

```

```

Db 218 TFNLNLNDILNSRLKRRKYFLDVLESFLMQFKHISSEYIIEDSFKLLNSEQKNTLLKSY 277
QY 159 KYIKESVENDIRFAQEGISYYEKVLAKYKDDLESIKKVIKEEKFPSPPTPPSPAKT 218
Db 278 KYIKESVENDIRFAQEGISYYEKVLAKYKDDLESIKKVIKEEKFPSPPTPPSPAKT 337
QY 219 DEQKESKFLPFLTNTETLYNNLVNKIDDDYLNLKAKINDCNVKEDEAHVKITKLSDLKA 278
Db 338 DEQKESKFLPFLTNTETLYNNLVNKIDDDYLNLKAKINDCNVKEDEAHVKITKLSDLKA 397
QY 279 IDDKIDLFRN 288
Db 398 IDDKIDLFRN 407

RESULT 10
Q25721 ID Q25721 PRELIMINARY; PRT; 373 AA.
AC Q25721;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1, 42 kDa C-terminal region (Fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-INFECTED HUMAN BLOOD SAMPLE;
RA Shi Y.-P., Alpers M.P., Povoa M.M., Nahlen B.L., Oloo A.G., Lal A.A.;
RT "Sequence of the C-terminal region of merozoite surface protein-1
RT (MSP-1) in field-derived Plasmodium falciparum.";
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U20726; AAG62217.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
FT NON_TER 1
FT NON_TER 373
SQ SEQUENCE 373 AA; 42848 MW; EE0700233D7F4D4E CRC64;

Query Match 43.9%; Score 189; DB 5; Length 373;
Best Local Similarity 100.0%; Pred. No. 2.2e-186;
Matches 189; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 AISTVMDNITLSPENEDVYILKPLAGVYRSLKKQIEKNIFTNLNLNDILNSRLKRRKY 117
Db 1 AISTVMDNITLSPENEDVYILKPLAGVYRSLKKQIEKNIFTNLNLNDILNSRLKRRKY 60

QY 118 FLDVLESFLMQFKHISSEYIIEDSFKLLNSEQKNTLLKSYKIKESVENDIRFAQEGIS 177
Db 61 FLDVLESFLMQFKHISSEYIIEDSFKLLNSEQKNTLLKSYKIKESVENDIRFAQEGIS 120

QY 178 YVEKVLAKYKDDLESIKKVIKEEKFPSPPTPPSPAKTDEQKESKFLPFLTNTETL 237
Db 121 YVEKVLAKYKDDLESIKKVIKEEKFPSPPTPPSPAKTDEQKESKFLPFLTNTETL 180

QY 238 YNNLVNKID 246
Db 181 YNNLVNKID 189

RESULT 11
Q25727 ID Q25727 PRELIMINARY; PRT; 373 AA.
AC Q25727;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1, 42 kDa C-terminal region (Fragment).
GN MSP-1.
OS Plasmodium falciparum.

```

```
OC Eukaryote; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=INFECTED HUMAN BLOOD SAMPLE;
RA Shi Y.-P., Alpers M.P., Povoa M.M., Nahlen B.L., Oloo A.G., Lal A.A.;
RT "Sequence of the C-terminal region of merozoite surface protein-1
(MSP-1) in field-derived Plasmodium falciparum."
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U20732; AAA62223.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
FT NON_TER 1 373
FT SEQUENCE 373 AA; 42814 MW; A799666CF38CC405C CRC64;

Query Match 38.3%; Score 165; DB 5; Length 373;
Best Local Similarity 100.0%; Pred. No. 2.4e-163;
Matches 165; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 TFNLNLNDILNSRLKRRKRYFLDVLESQKHISSNEYIIEFSKLLNSEQKNTLLKSY 158
Db 42 TFNLNLNDILNSRLKRRKRYFLDVLESQKHISSNEYIIEFSKLLNSEQKNTLLKSY 101

QY 159 KYIKESVENDIKFAQEGISYEKVLAKYKDDLESIKKVIKEEKFPSSPTTPPSPAKT 218
Db 102 KYIKESVENDIKFAQEGISYEKVLAKYKDDLESIKKVIKEEKFPSSPTTPPSPAKT 161

QY 219 DEQKESKFLPFLTNIETLYNNLVNKIDYILNLKAKINDCNVEK 263
Db 162 DEQKESKFLPFLTNIETLYNNLVNKIDYILNLKAKINDCNVEK 206

RESULT 12
Q25728 PRELIMINARY; PRT; 373 AA.
AC Q25728;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1, 42 kDa C-terminal region (Fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=INFECTED HUMAN BLOOD SAMPLE;
RA Shi Y.-P., Alpers M.P., Povoa M.M., Nahlen B.L., Oloo A.G., Lal A.A.;
RT "Sequence of the C-terminal region of merozoite surface protein-1
(MSP-1) in field-derived Plasmodium falciparum."
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U20733; AAA62224.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
FT NON_TER 1 373
FT SEQUENCE 373 AA; 42815 MW; 9CF4DDAF38CC4054 CRC64;

Query Match 38.3%; Score 165; DB 5; Length 373;
Best Local Similarity 100.0%; Pred. No. 2.4e-163;
Matches 165; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 TFNLNLNDILNSRLKRRKRYFLDVLESQKHISSNEYIIEFSKLLNSEQKNTLLKSY 158
Db 42 TFNLNLNDILNSRLKRRKRYFLDVLESQKHISSNEYIIEFSKLLNSEQKNTLLKSY 101

QY 159 KYIKESVENDIKFAQEGISYEKVLAKYKDDLESIKKVIKEEKFPSSPTTPPSPAKT 218
Db 102 KYIKESVENDIKFAQEGISYEKVLAKYKDDLESIKKVIKEEKFPSSPTTPPSPAKT 161

QY 219 DEQKESKFLPFLTNIETLYNNLVNKIDYILNLKAKINDCNVEK 263
Db 162 DEQKESKFLPFLTNIETLYNNLVNKIDYILNLKAKINDCNVEK 206
```

```
Db 162 DEQKESKFLPFLTNIETLYNNLVNKIDYILNLKAKINDCNVEK 206

RESULT 13
Q43997 PRELIMINARY; PRT; 372 AA.
AC Q43997;
DT 01-JUN-1998 (TEMBLrel. 06, Created)
DT 01-JUN-1998 (TEMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE Merozoite surface protein-1 (Fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KENYA-3;
RA Qari S.H., Shi Y.P., Goldman I.F., Nahlen B.L., Tibayrenc M., Lal A.A.;
RT "Predicted and observed alleles of Plasmodium falciparum merozoite
surface protein-1 (MSP-1), a potential malaria vaccine antigen."
RL Mol. Biochem. Parasitol. 92:241-252(1998).
DR EMBL; AF040369; AAC39099.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
FT NON_TER 1 372
FT SEQUENCE 372 AA; 42768 MW; 3ACD3AFA6C047D53 CRC64;

Query Match 33.6%; Score 145; DB 5; Length 372;
Best Local Similarity 100.0%; Pred. No. 1.8e-142;
Matches 145; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 AISVTMDNLSGPFENYDVLYLPLAGVYRSLLKQIEKNIFTFNLNDILNSRLKRRKY 117
Db 1 AISVTMDNLSGPFENYDVLYLPLAGVYRSLLKQIEKNIFTFNLNDILNSRLKRRKY 60

QY 118 FLDVLESQKHISSNEYIIEFSKLLNSEQKNTLLKSYKESVENDIKFAQEGIS 177
Db 61 FLDVLESQKHISSNEYIIEFSKLLNSEQKNTLLKSYKESVENDIKFAQEGIS 120

QY 178 YEKVLAKYKDDLESIKKVIKEEKE 202
Db 121 YEKVLAKYKDDLESIKKVIKEEKE 145

RESULT 14
Q25725 PRELIMINARY; PRT; 372 AA.
AC Q25725;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1, 42 kDa C-terminal region (Fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=INFECTED HUMAN BLOOD SAMPLE;
RA Shi Y.-P., Alpers M.P., Povoa M.M., Nahlen B.L., Oloo A.G., Lal A.A.;
RT "Sequence of the C-terminal region of merozoite surface protein-1
(MSP-1) in field-derived Plasmodium falciparum."
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U20730; AAA62221.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
FT NON_TER 1 372
FT SEQUENCE 372 AA; 42711 MW; E3DBB33F282CD5FC CRC64;
```



```
Db 259 AISVTMDNLGSEFENEYDVIYKPLAGVYRSLSKQIEKNIFTFNLNLDIILNSRLKRRKY 318
Qy 118 FLDVLESLDMQFKHSSNEYIIEISFKLLNSEQKNTLLSKSYKIKESVENDIKFAQEGIS 177
Db 319 FLDVLESLDMQFKHSSNEYIIEISFKLLNSEQKNTLLSKSYKIKESVENDIKFAQEGIS 378
Qy 178 YEKVLAKYKDDLESIKKVIKEKE 202
Db 379 YEKVLAKYKDDLESIKKVIKEKE 403

RESULT 18
Q25970
ID Q25970 PRELIMINARY; PRT; 569 AA.
AC Q25970;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13347; BAA02608.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
DR EGF-like domain; Merozoite.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64502 MW; 4D761FF472753142 CRC64;

Query Match 32.7%; Score 141; DB 5; Length 569;
Best Local Similarity 100.0%; Pred. NO. 4e-138;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 291 DFEAIKKLINDTTKKMLGKLLSTGLVQNFNTIISKLEIEGFQDMLNISQHCYKQCP 350
Db 409 DFEAIKKLINDTTKKMLGKLLSTGLVQNFNTIISKLEIEGFQDMLNISQHCYKQCP 468
Qy 351 ENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENGGCDADATCTEEDSGSSRRK 410
Db 469 ENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENGGCDADATCTEEDSGSSRRK 528
Qy 411 ITCECTKPDSPYPLFDGIFCSS 431
Db 529 ITCECTKPDSPYPLFDGIFCSS 549

Query Match 32.7%; Score 141; DB 5; Length 569;
Best Local Similarity 100.0%; Pred. NO. 4e-138;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 291 DFEAIKKLINDTTKKMLGKLLSTGLVQNFNTIISKLEIEGFQDMLNISQHCYKQCP 350
Db 409 DFEAIKKLINDTTKKMLGKLLSTGLVQNFNTIISKLEIEGFQDMLNISQHCYKQCP 468
Qy 351 ENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENGGCDADATCTEEDSGSSRRK 410
Db 469 ENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENGGCDADATCTEEDSGSSRRK 528
Qy 411 ITCECTKPDSPYPLFDGIFCSS 431
Db 529 ITCECTKPDSPYPLFDGIFCSS 549

RESULT 19
Q25980
ID Q25980 PRELIMINARY; PRT; 569 AA.
AC Q25980;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13354; BAA02615.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
DR EGF-like domain; Merozoite.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64530 MW; 4D734580516D2142 CRC64;

Query Match 32.7%; Score 141; DB 5; Length 569;
Best Local Similarity 100.0%; Pred. NO. 4e-138;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 291 DFEAIKKLINDTTKKMLGKLLSTGLVQNFNTIISKLEIEGFQDMLNISQHCYKQCP 350
Db 409 DFEAIKKLINDTTKKMLGKLLSTGLVQNFNTIISKLEIEGFQDMLNISQHCYKQCP 468
Qy 351 ENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENGGCDADATCTEEDSGSSRRK 410
Db 469 ENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENGGCDADATCTEEDSGSSRRK 528
Qy 411 ITCECTKPDSPYPLFDGIFCSS 431
Db 529 ITCECTKPDSPYPLFDGIFCSS 549

RESULT 21
Q25717
ID Q25717 PRELIMINARY; PRT; 372 AA.
```

```
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13353; BAA02614.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
DR EGF-like domain; Merozoite.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64536 MW; 8008861DBECCD8DC CRC64;

Query Match 32.7%; Score 141; DB 5; Length 569;
Best Local Similarity 100.0%; Pred. NO. 4e-138;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 291 DFEAIKKLINDTTKKMLGKLLSTGLVQNFNTIISKLEIEGFQDMLNISQHCYKQCP 350
Db 409 DFEAIKKLINDTTKKMLGKLLSTGLVQNFNTIISKLEIEGFQDMLNISQHCYKQCP 468
Qy 351 ENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENGGCDADATCTEEDSGSSRRK 410
Db 469 ENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENGGCDADATCTEEDSGSSRRK 528
Qy 411 ITCECTKPDSPYPLFDGIFCSS 431
Db 529 ITCECTKPDSPYPLFDGIFCSS 549

RESULT 20
Q25982
ID Q25982 PRELIMINARY; PRT; 569 AA.
AC Q25982;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13354; BAA02615.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
DR EGF-like domain; Merozoite.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64530 MW; 4D734580516D2142 CRC64;

Query Match 32.7%; Score 141; DB 5; Length 569;
Best Local Similarity 100.0%; Pred. NO. 4e-138;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 291 DFEAIKKLINDTTKKMLGKLLSTGLVQNFNTIISKLEIEGFQDMLNISQHCYKQCP 350
Db 409 DFEAIKKLINDTTKKMLGKLLSTGLVQNFNTIISKLEIEGFQDMLNISQHCYKQCP 468
Qy 351 ENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENGGCDADATCTEEDSGSSRRK 410
Db 469 ENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENGGCDADATCTEEDSGSSRRK 528
Qy 411 ITCECTKPDSPYPLFDGIFCSS 431
Db 529 ITCECTKPDSPYPLFDGIFCSS 549

RESULT 21
Q25717
ID Q25717 PRELIMINARY; PRT; 372 AA.
```



```
AC Q25717;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1, 42 kDa C-terminal region (Fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=INFECTED HUMAN BLOOD SAMPLE;
RA Shi Y.-P., Alpers M.P., Povoa M.M., Nahlen B.L., Oloo A.G., Lal A.A.;
RT "Sequence of the C-terminal region of merozoite surface protein-1
  (MSP-1) in field-derived Plasmodium falciparum.";
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U20653; AAA62213.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
FT NON_TER 1
FT NON_TER 372
SQ SEQUENCE 372 AA; 42687 MW; 3F2BF1152598FB10 CRC64;

Query Match 32.5%; Score 140; DB 5; Length 372;
Best Local Similarity 100.0%; Pred. No. 3e-137;
Matches 140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 DFEAIKKLLNDTTPKDMGLKLLSTGLVQNPNTIISKLEGKFDQMLNISQHCVKKQCP 350
DB 233 DFEAIKKLLNDTTPKDMGLKLLSTGLVQNPNTIISKLEGKFDQMLNISQHCVKKQCP 292

QY 351 ENSGCFRHLDERECKCLLNKQEGDKVCVENPNPTCNENNGGCDADATCTEEDSGSSRKK 410
DB 293 ENSGCFRHLDERECKCLLNKQEGDKVCVENPNPTCNENNGGCDADATCTEEDSGSSRKK 352

QY 411 ITCECTKPDSPYPLFDGIFCS 430
DB 353 ITCECTKPDSPYPLFDGIFCS 372

Query Match 32.5%; Score 140; DB 5; Length 372;
Best Local Similarity 100.0%; Pred. No. 3e-137;
Matches 140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 DFEAIKKLLNDTTPKDMGLKLLSTGLVQNPNTIISKLEGKFDQMLNISQHCVKKQCP 350
DB 233 DFEAIKKLLNDTTPKDMGLKLLSTGLVQNPNTIISKLEGKFDQMLNISQHCVKKQCP 292

QY 351 ENSGCFRHLDERECKCLLNKQEGDKVCVENPNPTCNENNGGCDADATCTEEDSGSSRKK 410
DB 293 ENSGCFRHLDERECKCLLNKQEGDKVCVENPNPTCNENNGGCDADATCTEEDSGSSRKK 352

QY 411 ITCECTKPDSPYPLFDGIFCS 430
DB 353 ITCECTKPDSPYPLFDGIFCS 372

Query Match 32.5%; Score 140; DB 5; Length 372;
Best Local Similarity 100.0%; Pred. No. 3e-137;
Matches 140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 DFEAIKKLLNDTTPKDMGLKLLSTGLVQNPNTIISKLEGKFDQMLNISQHCVKKQCP 350
DB 233 DFEAIKKLLNDTTPKDMGLKLLSTGLVQNPNTIISKLEGKFDQMLNISQHCVKKQCP 292

QY 351 ENSGCFRHLDERECKCLLNKQEGDKVCVENPNPTCNENNGGCDADATCTEEDSGSSRKK 410
DB 293 ENSGCFRHLDERECKCLLNKQEGDKVCVENPNPTCNENNGGCDADATCTEEDSGSSRKK 352

QY 411 ITCECTKPDSPYPLFDGIFCS 430
DB 353 ITCECTKPDSPYPLFDGIFCS 372

Query Match 32.5%; Score 140; DB 5; Length 372;
Best Local Similarity 100.0%; Pred. No. 3e-137;
Matches 140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 DFEAIKKLLNDTTPKDMGLKLLSTGLVQNPNTIISKLEGKFDQMLNISQHCVKKQCP 350
DB 233 DFEAIKKLLNDTTPKDMGLKLLSTGLVQNPNTIISKLEGKFDQMLNISQHCVKKQCP 292
```

```
QY 351 ENSGCFRHLDERECKCLLNKQEGDKVCVENPNPTCNENNGGCDADATCTEEDSGSSRKK 410
DB 293 ENSGCFRHLDERECKCLLNKQEGDKVCVENPNPTCNENNGGCDADATCTEEDSGSSRKK 352

QY 411 ITCECTKPDSPYPLFDGIFCS 430
DB 353 ITCECTKPDSPYPLFDGIFCS 372

RESULT 23
Q25720 PRELIMINARY; PRT; 372 AA.
AC Q25720;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1, 42 kDa C-terminal region (Fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=INFECTED HUMAN BLOOD SAMPLE;
RA Shi Y.-P., Alpers M.P., Povoa M.M., Nahlen B.L., Oloo A.G., Lal A.A.;
RT "Sequence of the C-terminal region of merozoite surface protein-1
  (MSP-1) in field-derived Plasmodium falciparum.";
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U20656; AAA62216.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
FT NON_TER 1
FT NON_TER 372
SQ SEQUENCE 372 AA; 42686 MW; 583A33972DB05FB4 CRC64;

Query Match 32.5%; Score 140; DB 5; Length 372;
Best Local Similarity 100.0%; Pred. No. 3e-137;
Matches 140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 DFEAIKKLLNDTTPKDMGLKLLSTGLVQNPNTIISKLEGKFDQMLNISQHCVKKQCP 350
DB 233 DFEAIKKLLNDTTPKDMGLKLLSTGLVQNPNTIISKLEGKFDQMLNISQHCVKKQCP 292

QY 351 ENSGCFRHLDERECKCLLNKQEGDKVCVENPNPTCNENNGGCDADATCTEEDSGSSRKK 410
DB 293 ENSGCFRHLDERECKCLLNKQEGDKVCVENPNPTCNENNGGCDADATCTEEDSGSSRKK 352

QY 411 ITCECTKPDSPYPLFDGIFCS 430
DB 353 ITCECTKPDSPYPLFDGIFCS 372

Query Match 32.5%; Score 140; DB 5; Length 372;
Best Local Similarity 100.0%; Pred. No. 3e-137;
Matches 140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 DFEAIKKLLNDTTPKDMGLKLLSTGLVQNPNTIISKLEGKFDQMLNISQHCVKKQCP 350
DB 233 DFEAIKKLLNDTTPKDMGLKLLSTGLVQNPNTIISKLEGKFDQMLNISQHCVKKQCP 292

QY 351 ENSGCFRHLDERECKCLLNKQEGDKVCVENPNPTCNENNGGCDADATCTEEDSGSSRKK 410
DB 293 ENSGCFRHLDERECKCLLNKQEGDKVCVENPNPTCNENNGGCDADATCTEEDSGSSRKK 352

QY 411 ITCECTKPDSPYPLFDGIFCS 430
DB 353 ITCECTKPDSPYPLFDGIFCS 372

Query Match 32.5%; Score 140; DB 5; Length 372;
Best Local Similarity 100.0%; Pred. No. 3e-137;
Matches 140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 DFEAIKKLLNDTTPKDMGLKLLSTGLVQNPNTIISKLEGKFDQMLNISQHCVKKQCP 350
DB 233 DFEAIKKLLNDTTPKDMGLKLLSTGLVQNPNTIISKLEGKFDQMLNISQHCVKKQCP 292

QY 351 ENSGCFRHLDERECKCLLNKQEGDKVCVENPNPTCNENNGGCDADATCTEEDSGSSRKK 410
DB 293 ENSGCFRHLDERECKCLLNKQEGDKVCVENPNPTCNENNGGCDADATCTEEDSGSSRKK 352

QY 411 ITCECTKPDSPYPLFDGIFCS 430
DB 353 ITCECTKPDSPYPLFDGIFCS 372

Query Match 32.5%; Score 140; DB 5; Length 372;
Best Local Similarity 100.0%; Pred. No. 3e-137;
Matches 140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 DFEAIKKLLNDTTPKDMGLKLLSTGLVQNPNTIISKLEGKFDQMLNISQHCVKKQCP 350
DB 233 DFEAIKKLLNDTTPKDMGLKLLSTGLVQNPNTIISKLEGKFDQMLNISQHCVKKQCP 292
```

```

DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
FT NON_TER 1
FT NON_TER 372 372
SQ SEQUENCE 372 AA; 42734 MW; 19844F60D7011E52 CRC64;

Query Match 31.8%; Score 137; DB 5; Length 372;
Best Local Similarity 100.0%; Pred. No. 4.1e-134;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 ILSGFENEYDVIYKPLAGVYRSUKKQIEKNIFTFNLNLDILNSRLKRRKYFLDVLSD 125
DB 9 ILSGFENEYDVIYKPLAGVYRSUKKQIEKNIFTFNLNLDILNSRLKRRKYFLDVLSD 68
QY 126 LMQPKHSSNEYIIEDSFKLNSQKNTLKSRYIKESVENDIKFAQEGISYIEKVLAK 185
DB 69 LMQPKHSSNEYIIEDSFKLNSQKNTLKSRYIKESVENDIKFAQEGISYIEKVLAK 128
QY 186 YKDDLESIKKVIKEKE 202
DB 129 YKDDLESIKKVIKEKE 145

RESULT 25
Q9TVG8 PRELIMINARY; PRT; 218 AA.
AC Q9TVG8;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE MSAL protein (Fragment).
GN MSAL.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FIJ4, AND FIS1;
RX MEDLINE=99263453; PubMed=10329360;
RA Lalitha P.V., Malhotra P., Chattopadhyay, Chauhan V.S.;
RT "Plasmodium falciparum: Variations in the C-Terminal Cysteine-Rich
Region of the Merozoite Surface Protein-1 in Field Samples among
Indian Isolates.";
RL Exp. Parasitol. 92:12-18(1999).
DR EMBL; Y10600; CAA71609.1; -.
DR EMBL; Y10599; CAA71608.1; -.
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
FT NON_TER 1
FT NON_TER 218 218
SQ SEQUENCE 218 AA; 24558 MW; EBD6A08FED6290F0 CRC64;

Query Match 27.1%; Score 117; DB 5; Length 218;
Best Local Similarity 100.0%; Pred. No. 1.9e-113;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 DFEAIKKLINDTTKKMLGKLLSTGLVQNFNTTIISKLEGFQDMLNISQHCVKKQCP 350
DB 58 DFEAIKKLINDTTKKMLGKLLSTGLVQNFNTTIISKLEGFQDMLNISQHCVKKQCP 117
QY 351 ENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGCCDADA 407
DB 118 ENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGCCDADA 174

RESULT 26
Q25969 PRELIMINARY; PRT; 569 AA.
AC Q25969;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DE 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).

```

```

GN MSPL.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
major merozoite surface proteins (MSPL) of Plasmodium falciparum from
field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13346; BAA02607.1; -.
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
FT NON_TER 569 569
SQ SEQUENCE 569 AA; 64485 MW; A26C45914AF01396 CRC64;

Query Match 24.8%; Score 107; DB 5; Length 569;
Best Local Similarity 100.0%; Pred. No. 1.2e-102;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 DFEAIKKLINDTTKKMLGKLLSTGLVQNFNTTIISKLEGFQDMLNISQHCVKKQCP 350
DB 409 DFEAIKKLINDTTKKMLGKLLSTGLVQNFNTTIISKLEGFQDMLNISQHCVKKQCP 468
QY 351 ENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGCCDADA 397
DB 469 ENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGCCDADA 515

RESULT 27
Q25974 PRELIMINARY; PRT; 569 AA.
AC Q25974;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSPL.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
major merozoite surface proteins (MSPL) of Plasmodium falciparum from
field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13346; BAA02609.1; -.
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
FT NON_TER 569 569
SQ SEQUENCE 569 AA; 64457 MW; A2691FE569E80396 CRC64;

Query Match 24.8%; Score 107; DB 5; Length 569;
Best Local Similarity 100.0%; Pred. No. 1.2e-102;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 DFEAIKKLINDTTKKMLGKLLSTGLVQNFNTTIISKLEGFQDMLNISQHCVKKQCP 350
DB 409 DFEAIKKLINDTTKKMLGKLLSTGLVQNFNTTIISKLEGFQDMLNISQHCVKKQCP 468
QY 351 ENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGCCDADA 397
DB 469 ENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGCCDADA 515

RESULT 28

```

```
Q25975
ID Q25975 PRELIMINARY; PRT; 569 AA.
AC Q25975;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13349; BAA02610.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64459 MW; ED4284B2867C9703 CRC64;

Query Match 24.8%; Score 107; DB 5; Length 569;
Best Local Similarity 100.0%; Pred. No. 1.2e-102;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 DFEATKKLNDTCKDMLGKLLSTGLVQNFPTTIISKLEGGFQDMLNLSQHCVKKQCP 350
Db 409 DFEATKKLNDTCKDMLGKLLSTGLVQNFPTTIISKLEGGFQDMLNLSQHCVKKQCP 468

QY 351 ENSGCFRHLDERECKCLLNKQEGDKVCVENPNTCNENNGGCDADA 397
Db 469 ENSGCFRHLDERECKCLLNKQEGDKVCVENPNTCNENNGGCDADA 515

RESULT 29
Q25977
ID Q25977 PRELIMINARY; PRT; 569 AA.
AC Q25977;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13350; BAA02611.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64485 MW; A26C45914AF01396 CRC64;

Query Match 24.8%; Score 107; DB 5; Length 569;
Best Local Similarity 100.0%; Pred. No. 1.2e-102;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 DFEATKKLNDTCKDMLGKLLSTGLVQNFPTTIISKLEGGFQDMLNLSQHCVKKQCP 350
Db 409 DFEATKKLNDTCKDMLGKLLSTGLVQNFPTTIISKLEGGFQDMLNLSQHCVKKQCP 468
```

```
QY 351 ENSGCFRHLDERECKCLLNKQEGDKVCVENPNTCNENNGGCDADA 397
Db 469 ENSGCFRHLDERECKCLLNKQEGDKVCVENPNTCNENNGGCDADA 515

RESULT 30
Q25979
ID Q25979 PRELIMINARY; PRT; 569 AA.
AC Q25979;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13352; BAA02613.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64457 MW; A2691FE569E80396 CRC64;

Query Match 24.8%; Score 107; DB 5; Length 569;
Best Local Similarity 100.0%; Pred. No. 1.2e-102;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 DFEATKKLNDTCKDMLGKLLSTGLVQNFPTTIISKLEGGFQDMLNLSQHCVKKQCP 350
Db 409 DFEATKKLNDTCKDMLGKLLSTGLVQNFPTTIISKLEGGFQDMLNLSQHCVKKQCP 468

QY 351 ENSGCFRHLDERECKCLLNKQEGDKVCVENPNTCNENNGGCDADA 397
Db 469 ENSGCFRHLDERECKCLLNKQEGDKVCVENPNTCNENNGGCDADA 515

RESULT 31
Q25981
ID Q25981 PRELIMINARY; PRT; 1694 AA.
AC Q25981;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Major merozoite surface antigen.
GN GPI95.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FCCL/HN;
RA Shan Z.X., Yu X.B., Li X.R., Ma C.L., Fang J.M.;
RT "Molecular cloning and sequence analysis of major merozoite surface
RT antigen(gpi95)gene of Plasmodium falciparum isolate FCCL/HN."
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF218248; AAF27526.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW Merozoite.
SQ SEQUENCE 1694 AA; 192767 MW; B51634A49E0F6728 CRC64;

Query Match 24.8%; Score 107; DB 5; Length 1694;
Best Local Similarity 100.0%; Pred. No. 3.3e-102;
```

Query Match 24.8%; Score 107; DB 5; Length 1704;

RX	MEDLINE=21349158; PubMed=11456319;
RA	Da Silveira L.A., Ribeiro W.L., Kirchgatter K., Wunderlich G.,
RA	Matsuoka H., Tanabe K., Ferreira M.U.;
RT	"Sequence diversity and linkage disequilibrium within the merozoite
RT	surface protein-1 (MSP-1) locus of Plasmodium falciparum: a
RT	longitudinal study in Brazil.";
RL	J. Eukaryot. Microbiol. 48:433-439(2001).
DR	EMBL; AF290875; AAK82766.1; -.
DR	InterPro; IPR000561; EGF-like.
DR	Pfam; PF00008; EGF; 1
FT	NON_TER 1 95
FT	NON_TER 95 95
SQ	SEQUENCE 95 AA; 10668 MW; C7786C3BEDA322ED CRC64;
Query Match 20.2%; Score 87; DB 5; Length 95;	
Best Local Similarity 100.0%; Pred. No. 1.9e-82;	
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy	336 MLNISQHCVKVKQPENSGCFRHLDEREECKLLNYKQEGDKVCVENPNPTCNENGGCA 395       Db 1 MLNISQHCVKVKQPENSGCFRHLDEREECKLLNYKQEGDKVCVENPNPTCNENGGCA 60
Qy	396 DATCEEDSGSRKKITCECTKPDSYP 422       Db 61 DATCEEDSGSRKKITCECTKPDSYP 87
RESULT 38	
Q9BUJ5	
ID	Q9BJU5 PRELIMINARY; PRT; 77 AA.
AC	Q9BJU5;
DT	01-JUN-2001 (TrEMBLrel. 17, Created)
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE	Merozoite surface protein 1 (Fragment).
OS	Plasmodium falciparum.
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX	NCBI_TaxID=5833;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Eisen D.P., Cloonan N., Baddeley A., Erl R., Saul A.;
RT	"Antigenic drift and immune selection acting on merozoite surface
RT	protein 1-19 and merozoite surface protein 2 in independent field
RT	isolates of Plasmodium falciparum."
RL	Submitted (DBC-2000) to the EMBL/GenBank/DDBJ databases.
DR	EMBL; AF329518; AAK19337.1; -.
FT	NON_TER 1 77
FT	NON_TER 77 77
SQ	SEQUENCE 77 AA; 8587 MW; 4EA799FF4B4E69D4 CRC64;
Query Match 17.9%; Score 77; DB 5; Length 77;	
Best Local Similarity 100.0%; Pred. No. 4.3e-72;	
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy	349 CPENSGCFRHLDEREECKLLNYKQEGDKVCVENPNPTCNENGGCDATCTEEDSGSSR 408       Db 1 CPENSGCFRHLDEREECKLLNYKQEGDKVCVENPNPTCNENGGCDATCTEEDSGSSR 60
Qy	409 KKITCECTKPSYPLFD 425       Db 61 KKITCECTKPSYPLFD 77
RESULT 39	
Q9BUJ4	
ID	Q9BJU4 PRELIMINARY; PRT; 77 AA.
AC	Q9BJU4;
DT	01-JUN-2001 (TrEMBLrel. 17, Created)
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE	Merozoite surface protein 1 (Fragment).
OS	Plasmodium falciparum.
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

```
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RA Eisen D.P., Cloonan N., Baddeley A., Eri R., Saul A.;
RT "Antigenic drift and immune selection acting on merozoite surface
RT protein 1-19 and merozoite surface protein 2 in independent field
RT isolates of Plasmodium falciparum."
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF329519; AAK19338.1; -
FT NON_TER 1
FT NON_TER 77
SQ SEQUENCE 77 AA; 8587 MW; 4EA799FF4B4E69D4 CRC64;

Query Match 17.9%; Score 77; DB 5; Length 77;
Best Local Similarity 100.0%; Pred. No. 4.3e-72;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 349 CPENSGCFRHLDERECKCLLNKQEGDKVCVENPNTCNENNGGCDADATCTEEDSGSSR 408
Db 1 CPENSGCFRHLDERECKCLLNKQEGDKVCVENPNTCNENNGGCDADATCTEEDSGSSR 60

QY 409 KKITCECTKPDSPFLFD 425
Db 61 KKITCECTKPDSPFLFD 77

RESULT 40
Q9BJU3 PRELIMINARY; PRT; 77 AA.
AC Q9BJU3;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Merozoite surface protein 1 (Fragment).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RA Eisen D.P., Cloonan N., Baddeley A., Eri R., Saul A.;
RT "Antigenic drift and immune selection acting on merozoite surface
RT protein 1-19 and merozoite surface protein 2 in independent field
RT isolates of Plasmodium falciparum."
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF329520; AAK19339.1; -
FT NON_TER 1
FT NON_TER 77
SQ SEQUENCE 77 AA; 8587 MW; 4EA799FF4B4E69D4 CRC64;

Query Match 17.9%; Score 77; DB 5; Length 77;
Best Local Similarity 100.0%; Pred. No. 4.3e-72;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 349 CPENSGCFRHLDERECKCLLNKQEGDKVCVENPNTCNENNGGCDADATCTEEDSGSSR 408
Db 1 CPENSGCFRHLDERECKCLLNKQEGDKVCVENPNTCNENNGGCDADATCTEEDSGSSR 60

QY 409 KKITCECTKPDSPFLFD 425
Db 61 KKITCECTKPDSPFLFD 77

Search completed: May 12, 2003, 10:28:14
Job time : 68 secs
```